FILOZOFICKÁ FAKULTA
UNIVERZITY KARLOVY
V PRAZE

Fonetický ústav

DIPLOMOVÁ PRÁCE

Bc. Dita Hývlová

Způsoby využití základní frekvence pro identifikaci mluvčích

Ways of exploiting fundamental frequency for speaker identification

Praha, 2015

Vedoucí práce: doc. Mgr. Radek Skarnitzl, Ph.D.

Poděkování

Ráda bych na tomto místě poděkovala Radku Skarnitzlovi za jeho výborné pedagogické působení a za to, že mi po celou dobu studia byl ve všem trpělivě nápomocen. Dále děkuji Filozofické fakultě UK za poskytnutí vnitřního grantu (VG 38), který umožnil výzkum, na základě něhož vznikla i tato diplomová práce.

Prohlašuji, že jsem diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 17. 8. 2015

podpis

Abstrakt

Předkládaná diplomová práce se zabývá forenzním využitím údajů o základní frekvenci, konkrétně jejích středních hodnot a ukazatelů variability. Mezi fonetiky zabývajícími se forenzní analýzou řeči je obecně známo, že F0 nemá příliš velký potenciál jako parametr využitelný při identifikaci mluvčích, protože podléhá mnoha vnějším faktorům (jako je momentální citové rozpoložení, okolní hluk, přenosový kanál nebo dokonce vlastní snaha maskovat svůj hlas), jež zapříčiňují vysokou intraindividuální variabilitu. Přesto však platí, že forenzní užití F0 skýtá i určité výhody, například snadnost extrakce jejích hodnot ze signálu a nižší ovlivnitelnost lexikálním obsahem – na rozdíl od vokalických formantů. V této práci zkoumáme nahrávky osmi mužských mluvčích pořízené ve dvou mluvních stylech (spontánním a čteném) a porovnáváme příslušné ukazatele stability i variability základní frekvence, které jsou pokud možno robustní vůči proměnlivým vnějším okolnostem: za střední hodnoty je to základní hladina a za deskriptory variability percentilové rozpětí. Kromě toho si všímáme řečových jevů, jako je třepená fonace, které jsou idiosynkratické a napomáhají rozlišitelnosti daného mluvčího od ostatních.

Klíčová slova: forenzní fonetika, identifikace mluvčího, základní frekvence, čeština

Abstract

The present Master's thesis deals with the forensic use of fundamental frequency characteristics, specifically with F0 mean values and indicators of variability. Phoneticians who specialise in the forensic analysis of speech generally believe that F0 does not hold much potential as a parameter useful for speaker identification, mainly because it is easily influenced by extrinsic factors (e.g. the speaker's emotional state, interfering noise, transmission channel or even the speaker's own effort to mask his voice), which cause high intra-individual variability. Despite these facts, however, the forensic use of F0 offers a number of advantages, namely straightforward extraction from the speech signal and lower susceptibility to varying lexical content – unlike, for example, vowel formants. This thesis investigates the recordings of 8 male speakers made in two different speech styles (spontaneous and read) and compares the respective indicators of F0 stability and variability, in particular those that are robust in varying external conditions: that is, the baseline for mean values and the 10.-90. percentile range for variability indicators. Apart from that, we take into account phenomena such as the creaky voice, which are idiosyncratic and contribute to easier speaker discrimination.

Key words: forensic phonetics, speaker identification, fundamental frequency, Czech

Obsah

1 Úvod

Forenzní fonetika je v současné době oblastí, v níž se dochází k zajímavým objevům s výrazným přesahem do praxe. Fonetika obecně má status vědy na průsečíku několika disciplín, humanitních i přírodních, a díky tomu je při její aplikaci možné zapojovat poznatky z lingvistiky, biologie, fyziky, statistiky i informačních technologií.

Forenzní zaměření fonetiky je záležitostí začínající obdobím po druhé světové válce. Od té doby se díky technologickému pokroku prudce zlepšily možnosti nahrávání a zkoumání řeči pro kriminalistické účely. Cílem fonetického bádání je nalézt akustický parametr řeči, který by ideálně pro každého jednotlivce nabýval jiné a snadno odlišitelné hodnoty, a zároveň byl detekovatelný již z velmi krátké nahrávky (Skarnitzl, 2014: 20). Jednou z domén řečového projevu, které fonetici podrobují analýze při tomto hledání, je základní složka hlasu, od níž se odvíjí všechny ostatní hlasové charakteristiky – základní frekvence.

Základní frekvence je na fonetickém poli důkladně popsána, protože její měření je poměrně jednoduché, například ve srovnání s vokalickými formanty. Mnozí odborníci zabývající se idiosynkrasií řečového chování k ní přistupují s určitou skepsí, jelikož často vykazuje známky intraindividuální variability, jíž dokáže mluvčí dosáhnout i úmyslně maskováním hlasu. Zároveň se k ní však fonetické studie vracejí, vždy když se v jiné oblasti objeví nový výzkumný pohled, který by mohla i u základní frekvence přinést zajímavé výsledky, jako je tomu v této práci například u srovnání dlouhodobých distribucí napříč mluvními styly.

Předkládaná diplomová práce si klade za cíl prozkoumat spolehlivost, s jakou o základní frekvenci vypovídají ukazatele středních hodnot i variability F0, a to jak ty déle používané a osvědčené, tak ty nověji objevené. Toto hledisko chce spojit s vytvořením foneticky relevantního závěru, který by v budoucnu mohl přispět ke zrychlení forenzní analýzy F0.

Struktura práce nabídne postupné uvedení do problematiky základní frekvence. V první, teoretické části zohledníme chronologii řečového chování a nejprve nastíníme produkci hlasu, tedy mechanismus dýchání, anatomii hrtanu a následně mechaniku fonace. K těmto oddílům budeme průběžně připojovat i foneticky relevantní komentáře. Další kapitola se bude zabývat akustickými aspekty F0, neboť po tvorbě hlasu a řeči následuje její přenos prostředím. Poté

připojíme několik poznámek k percepci F0, ale stěžejní část teoretické stati je úvod do druhů forenzních a akustických analýz F0.

Druhá, experimentální část nabídne studii provedenou na dvou druzích řečového materiálu od 8 mužských mluvčích, v níž se zaměříme zblízka na deskriptory F0 popisující její průměrné chování a variabilitu. Zároveň zhodnotíme, jak dobře vystihují řečové idiosynkrasie, jež pozorujeme u jednotlivých mluvčích jinými dostupnými způsoby – zejména pomocí poslechové analýzy a čtením dlouhodobých distribucí F0. Závěr této části čtenáře seznámí s nejdůležitějšími výstupy diplomové práce.

2 Produkce hlasu

2. 1 Plíce

Ke vzniku základní frekvence je zapotřebí stabilního výdechového proudu, až potom přichází na řadu přesná vzájemná součinnost nebo protipohyb svalů a vazů v hrtanu, které zajišťují nastavení pro vznik vibrace.

Činností nádechových mezižeberních svalů a bránice se roztahuje hrudní koš a zvětšuje hrudní dutina, aby následně došlo k nucenému vyrovnání sníženého tlaku vzduchu v plicích s tlakem vzduchu ve vnějším okolí. K obyčejnému výdechu zpravidla není potřeba aktivního zapojení výdechových mezižeberních svalů, protože se žeberní chrupavky a plicní tkáň díky své elasticitě samy snaží vrátit do původního stavu. Při řeči se však klade nárok na relativně stálý tlak pod hlasivkami (*subglotální* tlak), proto musejí výdechové mezižeberní svaly dynamicky sekundovat přirozenému elastickému smršťování plic. Výdech při fonaci tvoří až 90% celkového trvání respiračního cyklu a spotřeba vzduchu často přesahuje normální klidovou kapacitu (Estenne *et al*., 1990; citováno v Kreiman a Sidtis, 2011: 30). Jak se postupně kapacita vyčerpává, musejí se stáhnout i břišní svaly. Zároveň je však třeba neustále korigovat poměr subglotálního a supraglotálního tlaku, což je ztěžováno proměnlivým zužováním vokálního traktu při výslovnosti různých hlásek.

Načasování jednotlivých svalových aktivit při výdechu je tedy poměrně komplikovaný proces, který dosud nebyl ve fonetickém výzkumu plně probádán. Je možné, že se toto časování liší mluvčí od mluvčího. Idiosynkratické taktiky však již byly prokázány u přizpůsobování objemu plic pro hlasitější mluvu (Winkworth *et al.*, 1994; citováno v Kreiman a Sidtis, 2011: 31) nebo tendence k využívání spíše mezižeberních či spíše břišních svalů. Tyto návyky se vytvářejí krátce po narození (Boliek *et al*., 1996; citováno v Kreiman a Sidtis, 2011: 31).

2.2 Hrtan a jeho chrupavky

Základní funkcí hrtanu není tvorba hlasu, ale zamezení vniku cizorodých částic a těles do plic – zejména při polykání – a naopak vypuzení cizorodých částic ven vytvořením tlaku vzduchu při kašli. Pevným semknutím hlasivek po nádechu se také vytvoří opora pro zvedání

nebo tlačení těžkých předmětů. Také dýchání jako základní biologická funkce má před tvorbou hlasu přednost, proto když vyvíjíme náročnou pohybovou aktivitu, stará se hrtan hlavně o dostatečnou průchodnost vzduchu, a řeč je až na posledním místě.

Přesto se lidský hrtan vyvinul ve velice složitý systém, umožňující širokou paletu různých hlasových nastavení. Je tvořen soustavou chrupavek vzájemně spojených vazy a svaly. Je zavěšen pod jazylkou, která jako jediná kost v celém těle není kloubně spojena s jinou kostí či chrupavkou, pouze vazivem s horními rohy chrupavky štítné (latinsky *cartilago thyroidea*) a s příklopkou hrtanovou (*cartilago epiglottica*), která zastává důležitou funkci při ochranném uzavírání dýchacího ústrojí. Vrchní část příklopky hrtanové je připojena ke kořeni jazyka.

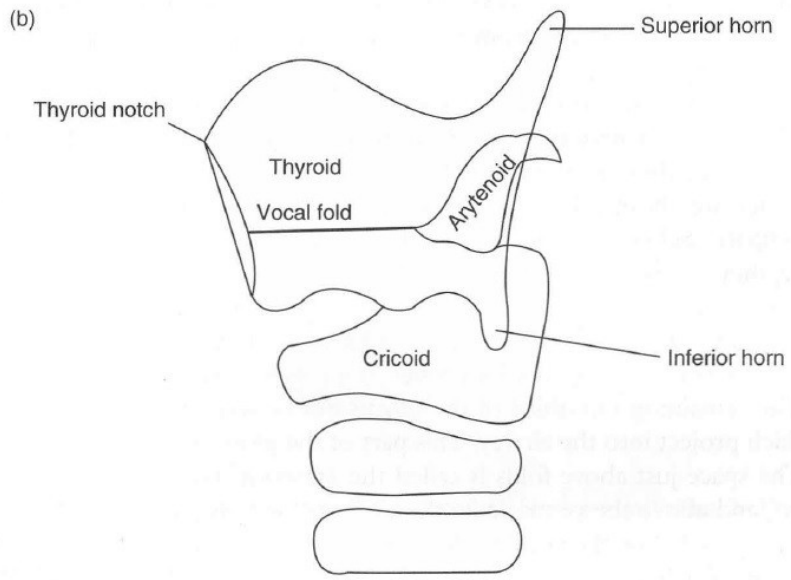
Za horní okraj samotného hrtanu můžeme považovat hranu chrupavky štítné, která vepředu dvěma ploténkami svírá ostrý úhel – u mužů ostřejší než u žen. Pro zajímavost můžeme uvést, že ačkoli je chrupavka štítná největší chrupavkou v hrtanovém systému, její rozměr od špičky horních rohů po špičky dolních rohů se pohybuje kolem pouhých 4 centimetrů.

Zespoda hrtan nasedá na průdušnici (*trachea*) prstencovou chrupavkou (*cartilago cricoidea*). Ta je ve své zadní části zesílená a vysoká. S prstencovou chrupavkou je štítná chrupavka kloubně spojená dolními rohy v zadní části. Tento kriko-thyroidní kloub je naprosto zásadní z hlediska regulace základní frekvence hlasu, protože umožňuje přibližování štítné chrupavky „stahováním“ její přední části směrem dolů k chrupavce prstencové. V důsledku tohoto pohybu dochází k napínání hlasivek (mechaniku fonace podrobně popíšeme v oddíle 2.4).

Na horní hraně zadní části prstencové chrupavky se nacházejí hlasivkové chrupavky (*cartilago arytaenoidea*), které mají tvar trojbokých pyramid a s prstencovou chrupavkou jsou spojeny výjimečně pohyblivým (kriko-arytenoidním) kloubem. Houpavé, klouzavé a rotační pohyby hlasivkovým chrupavkám umožňují širokou škálu postavení, ať už k různě usilovnému dýchání, šepotu nebo k samotné fonaci. Chrupavky se mohou přibližovat k sobě, vtáčet směrem dovnitř do hrtanu nebo mírně vytáčet po svých osách do stran, přičemž při řečové činnosti se kombinují všechny tyto pohyby (Seikel *et al*., 2010: 185).

Hlasivkové chrupavky mají obě ve své dolní přední části výběžek *processus vocalis*, ze kterého směrem ke štítné chrupavce vycházejí párové hlasivkové vazy a svaly. V dolní zadní

části najdeme výběžek *processus muscularis.* Na něj jsou rovněž navázány svaly, jejichž funkci upřesníme dále.



Obr. 2.1. Schematické znázornění hlavních hrtanových chrupavek z pohledu zboku, jejich výběžků a upnutí hlasivek. Pozn. autora: *thyroid – chrupavka štítná, thyroid notch – výběžek chrupavky štítné, vocal fold – hlasivka, arytenoid – chrupavka hlasivková, cricoid – chrupavka prstencová, superior/inferior horn – horní/dolní roh chrupavky štítné.* Převzato z Kreiman a Sidtis (2011).

2.3 Svaly hrtanu

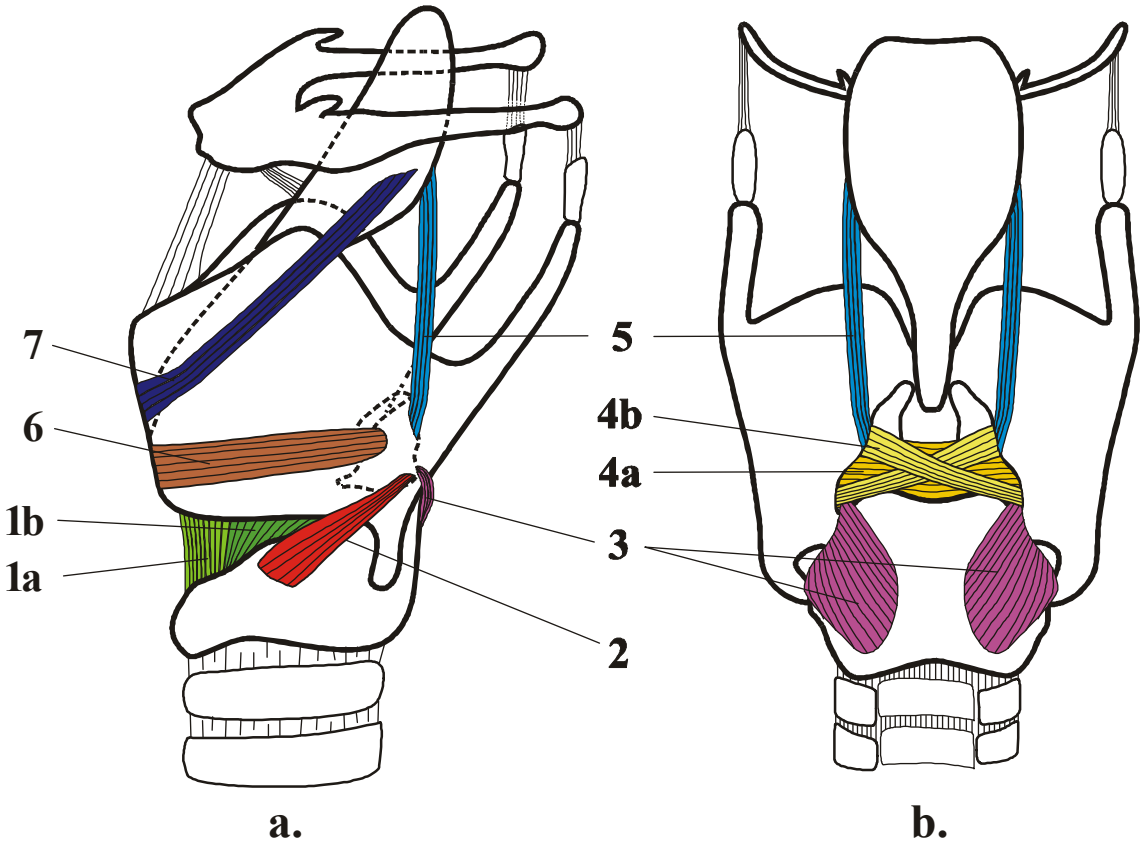
Svaly hrtanu dělíme na vnější a vnitřní. Vnější svaly upevňují hrtan v jeho zavěšení pod jazylkou a pohybují celým hrtanem nahoru a dolů. Vnitřní svaly mají na starost velmi rychlé a přesné změny v postavení hrtanových chrupavek vůči sobě, zejména chrupavek hlasivkových. Podle protikladných funkcí hovoříme u vnitřních hrtanových svalů o *adduktorech*, které k sobě hlasivkové chrupavky přibližují, a *abduktorech*, které je naopak oddalují a tím rozevírají hlasivkovou štěrbinu, neboli *glotis.* Dále v hrtanu existují svaly, které svou kontrakcí buď napínají, nebo uvolňují hlasivky, a tím regulují základní frekvenci. Na tomto místě je vhodné poznamenat, že činnost jedné nebo druhé skupiny svalů (adduktorů a abduktorů, napínačů a uvolňovačů) není vzájemně výlučná; tím, že kontrakce při současném zapojení působí proti sobě, dochází k jemné regulaci napětí v hrtanu a k ukotvení hlasivek.

Za hlavní **adduktory** považujeme laterální kriko-arytenoidní sval a příčný i šikmý arytenoidní sval (viz obr. 2.2). Umístění laterálního kriko-arytenoidního svalu není jednoduché si představit. Vychází z horních okrajů bočních stran chrupavky prstencové a upíná se na zadní, svalový výběžek chrupavek hlasivkových. Právě tento párový sval je zodpovědný za výše zmíněný vtáčivý pohyb hlasivkových chrupavek směrem dovnitř, kterým se k sobě přibližují přední výběžky hlasivkových chrupavek a tím i hlasivkové vazy. Kontrakce tohoto svalu však sama nezpůsobí úplné uzavření glotis, protože k sobě nejsou hlasivkové chrupavky pevně semknuty. Jejich vnitřní plochy definují celou jednu třetinu délky štěrbiny. Tato část se nazývá chrupavčitá glotis a zbytek, ohraničený hlasivkovými vazy, blanitá glotis.

Laterální kriko-arytenoidní sval svou činností tedy nastaví hlasivky na šepot. Chceme-li vytvořit podmínky k fonaci, musíme zapojit arytenoidní sval. Ten se nazývá také interarytenoidní, protože vzájemně propojuje hlasivkové chrupavky v jejich zadní části. Přibližuje je k sobě klouzavým pohybem po prstencové chrupavce a umožňuje také větší středové stlačení hlasivek při zvýšeném mluvním úsilí (Seikel *et al*., 2010: 191).

Jako hlavní **abduktor** funguje posteriorní kriko-arytenoidní sval, jehož činnost je protikladná k činnosti laterálního kriko-arytenoidního svalu. Propojuje (jak název vždy napovídá) zadní část prstencové chrupavky a svalové výběžky hlasivkových chrupavek, které při kontrakci stahuje dolů a tím rozevírá přední výběžky chrupavek, tedy glotis ve střední části, aby mohlo docházet ke snadnějšímu proudění vzduchu do plic. Tento sval je rovněž zapojován jako protiklad ke kriko-thyroidnímu svalu.

Kriko-thyroidní sval řadíme mezi hlasivkové napínače. Autoři Shipp a McGlone v roce 1971 zjistili, že kriko-thyroidní sval má nejdůležitější úlohu při regulaci F0. Spojuje štítnou a prstencovou chrupavku v jejich přední části a kontrakcí je přibližuje k sobě, tedy táhne štítnou chrupavku dolů. Štítná chrupavka se tímto pohybem oddálí od hlasivkových chrupavek, čímž se prodlouží a napnou hlasivkové vazy. Aby mohlo docházet k jemným úpravám napnutí hlasivkových vazů, je zapotřebí protitahu a ukotvení, které zajišťuje právě posteriorní kriko-arytenoidní sval. Dokonce je tak významný, že je samostatně inervován horním laryngálním nervem (zatímco zbytek hrtanového svalstva dolním laryngálním nervem). Inervace různými nervy umožňuje povolování jednoho svalu za současného stahování jiného, což je zásadní pro zmíněné antagonistické působení.



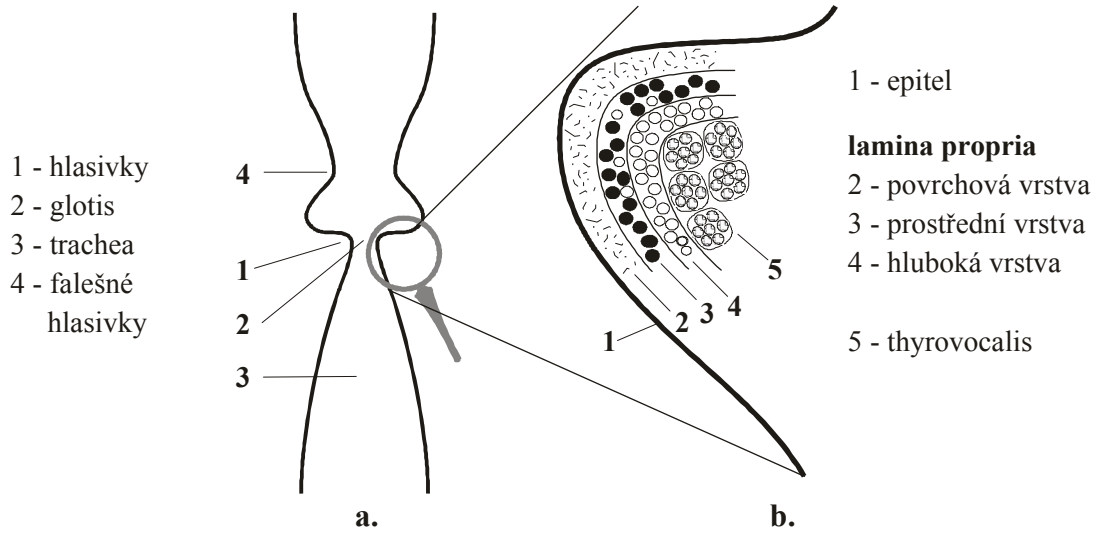
1a – kriko-thyroidní sval, zprava	4b – šikmý arytenoidní sval
1b – kriko-thyroidní sval, zleva	5 – ary-epiglotticus
2 – laterální kriko-arytenoidní sval	6 – thyro-arytenoidní sval
3 – posteriorní kriko-arytenoidní sval	7 – thyro-epiglotticus
4a – příčný arytenoidní sval	

Obr. 2.2. Hrtanové svalstvo. (a. – pohled zleva, b. – pohled zezadu). Svaly se vždy nazývají podle chrupavek, které spojují. Upraveno podle Skarnitzl (2011: 21).

Funkčním protikladem ke kriko-thyroidnímu svalu je také thyro-arytenoidní sval (přesněji jeho část *thyrovocalis*), který od sebe štítnou a prstencovou chrupavku oddaluje. Na tomto místě je vhodné přejít k popisu stavby samotných hlasivek, které se dělí do několika vrstev, odlišených jak fyziologicky, tak funkčně.

Již jsme hovořili o hlasivkových vazech, které jsou nataženy mezi předními výběžky hlasivkových chrupavek a vnitřním úhlem štítné chrupavky, směrem dopředu. Hlasivkové vazy jsou ale jen svrchní vrstvou masy hlasivek, které je třeba si představit jako vychlípení či

„ztluštění“ hrtanové stěny (viz obr. 2.3.). Anglický výraz *vocal folds* poměrně dobře vystihuje záhybový charakter hlasivek. Nad hlasivkami se nacházejí podobné, ale menší záhyby, výchlipkové řasy známé jako „falešné hlasivky“. S jejich pomocí se dá také tvořit jakýsi hlas, ale obtížně a jen v případech, kdy primární hlasivky nemohou plnit svoji funkci.



Obr. 2.3. Průřez hrtanem a průřez pravou hlasivkou. Převzato ze Skarnitzl (2011: 25).

Jak je zřejmé z obrázku 2.3 a jak uvádí i Seikel *et al*. (2010: 174), hlasivky jsou tvořeny pěti vrstvami tkáně, jejichž textura se směrem do hloubky mění podle potřebné funkce. Úplný povrch hlasivek tvoří velmi tenký epitel, který zachovává tvar hlasivek, chrání je před poškozením při jejich vzájemném nárazu a udržuje je zvlhčené. Pod tímto epitelem nacházíme tři vrstvy vaziva, souhrnně nazvané *lamina propria* (tloušťka je asi 2-4 mm). Povrchová vrstva obsahuje elastinová vlákna, tudíž je nejpružnější a nejsnadněji se rozkmitává. Prostřední vrstva obsahuje už i vlákna kolagenová, která naopak pružná nejsou a dodávají hlasivkovým vazům pevnost. Hluboká vrstva je hlavně složena z těchto vláken.

Hlouběji do jádra hlasivky se dostáváme k thyro-arytenoidnímu svalu. Ten tvoří většinu hmoty hlasivek. Někteří autoři – např. Titze (1994), Kreiman a Sidtis (2011) – hovoří o jediném svalu, jiní – např. Seikel *et al.* (2010), Gick *et al.* (2013) – o jeho rozdělení na *thyrovocalis* a *thyromuscularis. Thyrovocalis* leží pod hlasivkovým vazivem a vede od štítné chrupavky k přednímu výběžku hlasivkové chrupavky, *thyromuscularis* leží hlouběji a upíná

se ke svalovému výběžku hlasivkové chrupavky. Hlavní argument pro toto rozdělení je odlišný účinek při kontrakci těchto svalových vrstev. Zatímco *thyrovocalis* hlasivky napíná, obzvlášť když je zapojen současně s kriko-thyroidním svalem, *thyromuscularis* je uvolňuje. Jako spodní vrstva totiž smrštěním způsobí povolení napětí ve vrstvách blíže k povrchu.

Epitel a první dvě vrstvy hlasivkového vaziva považujeme za obal hlasivek, hluboké vazivo a thyro-arytenoidní sval považujeme za jádro. Schopnost jádra hlasivek mít napětí odlišné od hlasivkového obalu je základním předpokladem pro fonaci, jejíž princip představíme v následujícím oddíle.

Protože je tato práce zaměřena na možnosti rozpoznání identity mluvčího ze základní frekvence extrahované z řečové nahrávky, nabízí se otázka, které užitečné anatomické idiosynkrasie lze v nahrávce řeči detekovat. Například délka vokálního traktu nebo hlasivek dostatečnou typičnost pro jednotlivce nevykazuje, protože se jen velmi málo případů vzdaluje od populačního průměru. Stejně vzácné jsou případy, ve kterých je odlišení mluvčích od sebe umožněno díky patologickým zvláštnostem, jako jsou uzlíky na hlasivkách nebo asymetrické kmitání. Rovněž nemá význam uvažovat krátkodobé změny v anatomii způsobené například nastydnutím. Shrňme tedy, že na úrovni laryngální anatomie tedy o přílišném prostoru k idiosynkrasii uvažovat nelze, pouze ve smyslu neutrálních dlouhodobých nastavení vokálního traktu u jednotlivých mluvčích, na něž lze usuzovat po odečtení segmentálních změn (viz např. Laver, 2000; citováno v Mackenzie Beck, 2005).

2.4 Mechanika fonace

Již dávno byla vyvrácena představa, že k tomu, aby hlasivky mohly kmitat, je zapotřebí opakované svalové aktivity. Adduktory a abduktory, které jsme zevrubně popsali v předchozí kapitole, mají za úkol hlasivky nastavit jako překážku výdechovému proudu a po skončení fonace je od sebe opět oddálit pro volné dýchání. Během samotné fonace je třeba jen, aby pro danou artikulovanou hlásku zachovaly napětí nutné pro nepřerušovanou vibraci. Pojďme se blíže podívat na předpoklad vzniku fonace a průběh fonačního cyklu.

Nejprve musejí adduktory zajistit fonační postavení hlasivek tak, aby se hlasivkové chrupavky dotýkaly svými předními výběžky, ale aby nebyly pevně semknuté (pokud daný

řečový úsek nezačíná rázem). Rovněž kriko-thyroidní a thyro-arytenoidní sval musejí být v přesné souhře, aby nebyly hlasivky příliš napnuté svým prodloužením či zatnutím svalového jádra.

Hajime Hirose (2010: 141) uvádí čtyři podmínky, za kterých může být zahájena a udržena normální fonace:

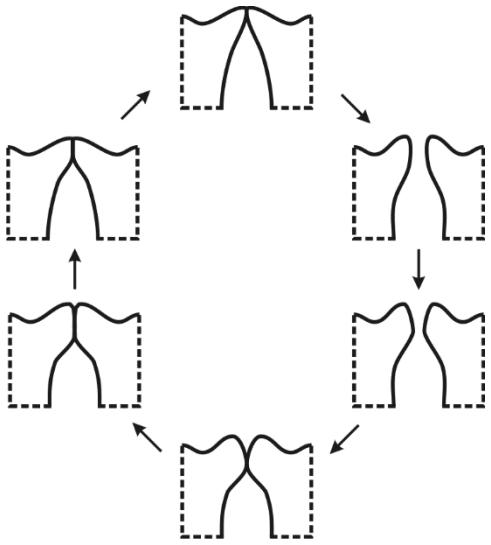
- dostatečná úroveň transglotálního tlaku (tzn. dostatečný rozdíl mezi subglotálním a supraglotálním tlakem vzduchu);
- dostatečně silný proud vzduchu vycházející z plic;
- hlasivky musí být relativně blízko u sebe;
- glotální odpor musí být dostatečně nízký.

Jak píše Radek Skarnitzl (2011: 39), všechny tyto parametry byly sloučeny do konceptu *fonační prahový tlak* (*phonation threshold pressure*), který definuje subglotální tlak potřebný k rozkmitání hlasivek a vystihuje snadnost fonace. Vnější faktory, které mají na fonační prahový tlak vliv, jsou např. hydratace hlasivek nebo F0, na které mají hlasivky kmitat; pro nižší F0 stačí tlak nižší a naopak.

Obal hlasivek, který jsme definovali jako epitel a první dvě vrstvy vaziva, představuje volně pohyblivou hmotu hlasivek, která má možnost se rozkmitat pod dostatečným tlakem vzduchu. Tlak, který je nutný pro rozražení hlasivek středově sevřených pro fonaci, se uvádí jako 3-5 H_2O. Po dosažení takového tlaku se hlasivky začnou rozevírat zespodu směrem nahoru, jak je patrné na obrázku 2.4, a stejným způsobem se odspodu začnou i zavírat, až je fonační cyklus ukončený. Sloupec vzduchu nad hlasivkami je uvedený do pohybu přerušovaným výdechovým proudem, a takto opakované excitace v molekulách vzduchu vytvoří zvukovou vlnu, která se může šířit prostředím.

K tomu, aby se kmitání hlasivek udrželo, stačí potom menší subglotální tlak, protože kmitání napomáhá několik fyzikálních aspektů. Předně je to elasticita kmitající tkáně, která se chce vrátit do původního tvaru. Dále je to Bernoulliho aerodynamický princip, podle nějž se výdechový proud zúžený hlasivkami zrychluje, objem molekul je v tomto místě řidší a klesá tím tlak mezi hlasivkami. Nižší tlak potom urychluje navrácení kontaktních ploch hlasivek k sobě. Poslední aerodynamickou silou jsou drobné víry, které vzniknou nad hlasivkami v momentě, kdy se opětovně uzavřou. Protože se část výdechového sloupce s vyšší hustotou

molekul posunula výše do vokálního traktu, v místě nad hlasivkami klesne tlak a tyto vzduchové víry vytvoří ještě lepší podmínky pro rychlé zavření hlasivek, což zesiluje energii vysokých frekvencí vznikajícího zvukového signálu a výsledkem je jasný hlas. U žen je normální, že se zadní část glotis nedovírá úplně, proto mívají hlas měkčí a dyšnější (Kreiman a Sidtis, 2011: 48). V každém případě platí, že zavírání hlasivek je prudší než jejich rozevírání a je to právě tato část cyklu, která způsobí největší výchylku tlaku, což se na zvukové vlně zobrazí jako ostrá špička v rámci jedné periody. Tyto výchylky tlaku molekul vzduchu potom doputují až k přijímajícímu sluchovému ústrojí, kde jsou převedeny na nervové vzruchy a mozkem zaznamenány jako zvuk.



Obr. 2.4. Schematické znázornění hlasivkového cyklu, pohled zepředu. Upraveno podle Skarnitzl (2011: 31).

Právě jsme popsali fonační cyklus tak, jak odpovídá *myoelasticko-aerodynamické* teorii tvorby hlasu, již navrhl roku 1958 Janwillem van den Berg. Ta byla vylepšena, když v roce 1974 Minoru Hirano představil svůj *body-cover* model, rozdělující hlasivky na jádro a obal (jak bylo popsáno výše). Každá z těchto částí se při kmitání chová jinak a velice záleží na jejich vzájemném napětí, když chceme ovládat základní frekvenci (F0). Ta se mění s druhou odmocninou poměru napětí : hmota (např. Titze, 1994: 193), zvýšením napětí tedy dosáhneme zvýšení F0. Se zvyšováním F0 nejsilněji koreluje aktivita kriko-thyroidního svalu, který natahuje hlasivky, a tím dosahuje většího napětí v jádru i obalu. Zároveň s ním se musí zapojit i posteriorní kriko-arytenoidní sval, aby ukotvil hlasivkové chrupavky, které by jinak byly taženy dopředu, a napnutí vazů by vyžadovalo větší námahu. Pokud se zároveň s kriko-

thyroidním svalem zapojí i *thyrovocalis,* může větším napětím přispět k dalšímu zvyšování F0. Za nižší aktivity kriko-thyroidního svalu se kontrakcí *thyrovocalis* hlasivky zkrátí a zmohutní. Zvětší se tak sice napětí jádra, ale zmenší se napětí obalu, tím pádem vznikne více efektivní hmoty (tedy hmoty činné v kmitání) a ta se může pohybovat s větší amplitudou, což by mělo být ideální nastavení pro střední a nižší F0. Jak poznamenává Titze (tamtéž: 194), chování výsledné základní frekvence za současného zapojení těchto svalů není s jistotou předvídatelné.

Je totiž obecně přijímáno, že zvýšením svalové aktivity se zvyšuje F0, ale zároveň platí, že vztah protažení hlasivek a napětí není přísně lineární, takže při zapojení kriko-thyroidního svalu může při prodlužování hlasivek na chvíli dojít ke snížení F0 (tamtéž: 201). Naopak je při zvyšování F0 – a to platí zejména pro zpěv – třeba postupně uvolňovat *thyrovocalis,* aby se napětí přeneslo na obal a zmenšila se tak efektivní kmitající hmota hlasivek. Je tedy patrné, že nezáleží ani tak na absolutní aktivitě jednoho nebo druhého svalu, ale spíše na rozdílu jejich aktivit a jak se tyto aktivity projevují na jednotlivých hlasivkových vrstvách.

Z výše popsaného vyplývá, že hlasivkové tkáně se pohybují nerovnoměrně. Obzvláště důležitý je nerovnoměrný pohyb hlasivek ve vertikálním směru (nezávislé otevírání/zavírání spodní a vrchní části hlasivek, viz obr. 2.4), protože střídáním sbíhavého a rozbíhavého tvaru glotis je umožněno samočinné kmitání hlasivek (Titze, 1986, 1994). Díky kombinaci této vertikální vlny a stojatého vlnění v horizontální rovině je hlasivkový tón bohatý na četné frekvenční složky, tedy harmonické složky.

Doposud jsme popisovali modální fonaci, tedy takovou, kterou lidské hlasivky provozují naprostou většinu času. Existují však i fonace nemodální, například falzet, používaný při zpěvu po přechodu do hlavového rejstříku, nebo takové fonační typy, které se někdy objeví v běžné řeči: tlačená, třepená a dyšná. Jsou to nepravidelnosti v už tak kvaziperiodickém hlasivkovém signálu, které jsou způsobeny nestandardními podmínkami fonace. V případě tlačené fonace je to neúměrně vysoký subglotální tlak za silného stlačení hlasivek. Takové nastavení hlasivky překrvuje a může velmi rychle poškodit. Třepená fonace je důsledkem naopak ubývajícího tlaku na konci dechového úseku (proto se s ní často setkáváme na konci promluv) – hlasivky zůstávají většinu cyklu zavřené. Dyšná fonace představuje přechod mezi normální fonací a šepotem, objevuje se při ní postupnější a ne tak ostré zavírání hlasivek, až

mohou hlasivky zůstávat dokonce mírně otevřené, díky čemuž se ke zvuku postupně přidá šum způsobený vzduchovou turbulencí.

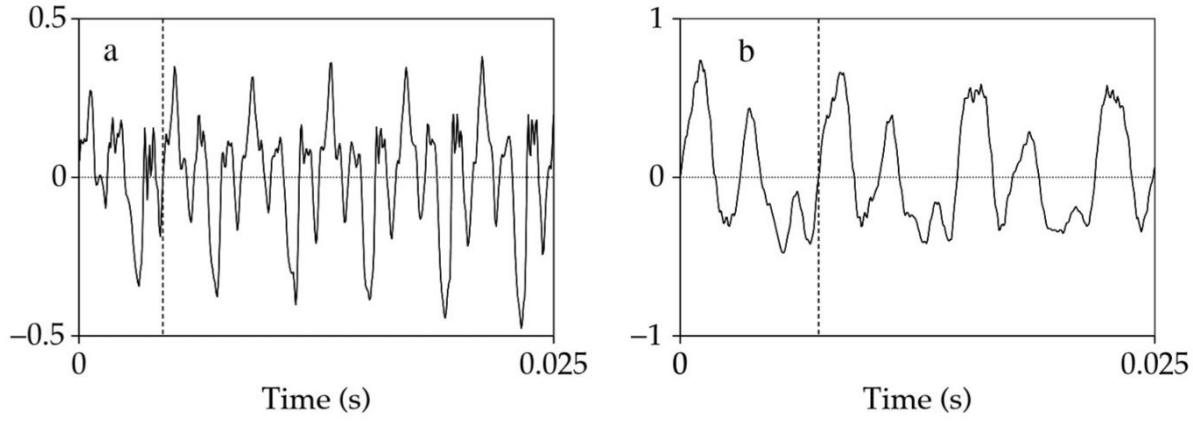
Závěrem tohoto oddílu se opět pokusme zasadit výše popsané do forenzního kontextu. Zvláštní typy fonace, které jsme vyjmenovali, můžeme považovat za idiosynkratické projevy řečového chování mluvčího, ke kterým je výhodné přihlédnout ve forenzní fonetice. Jejich výskyt není natolik vzácný, aby se na ně nedalo spolehnout v běžné forenzní praxi. Sklon k třepené fonaci se například u různých mluvčích vyskytuje v různých měrách. Nemodální fonaci řadíme k naučenému řečovému chování, do nějž dále patří obvyklá poloha F0, intonační kontury, ale také přízvuk a řečové tempo (Kreiman a Sidtis, 2011: 65). Musíme si ale také uvědomit, že kvalita hlasu nemusí být dána pouze fyziologicky nebo naučeným zacházením s hlasem, může se měnit i s denní dobou a současným fyzickým či psychickým rozpoložením (více v kapitole 6 o střední hlasové frekvenci a faktorech, které ji ovlivňují). Tyto faktory představují naopak nevýhodu pro forenzní praxi.

3 Akustické aspekty základní frekvence

3.1 Hlasivkový signál

Jak jsme si podrobně popsali, hlasivky opakováním fonačního cyklu vydávají rychle po sobě jdoucí pulsy, jež se slévají do zvuku o určité základní frekvenci. Čím rychleji hlasivky kmitají (ne náhodou se frekvenci říká také kmitočet), tím vyšší zvuk slyšíme. Na obrázku 3.1 vidíme záznam výchylek tlaku vzduchu kolem nuly, znázorňující průběh zvukové vlny v čase, tzv. oscilogram. Vlevo se jedná o ženský hlas, vpravo o mužský. Je zřejmé, že na stejném časovém úseku (25 ms) jsou kmity u ženského hlasu nahuštěnější. Rovněž lze pouhým okem poznat, že tvar zvukové vlny se v určitých intervalech opakuje. Tento interval se nazývá perioda, časově odpovídá jednomu fonačnímu cyklu a na obrázku je vyznačen svislou přerušovanou čarou. U ženy v obrázku trvá jedna perioda 4,42 ms a u muže 7,26 ms. Vydělením jedné vteřiny (1000 ms) periodou získáme základní frekvenci hlasu v jednotkách hertz (Hz), pro ženský hlas tedy v našem případě 226 Hz a pro mužský 138 Hz.

Jak jsme si výše uvedli, hlasivky jako živá tkáň nekmitají a ani nemohou kmitat tak, aby zvuková vlna měla tvar dokonalé sinusoidy, obrázek 3.1 tedy znázorňuje kvaziperiodický zvukový signál. Již jsme zmiňovali, že složitý způsob kmitání hlasivek obohacuje základní hlasivkový tón o další frekvenční složky. Přítomnost těchto složek je patrná na zubatém tvaru zvukové vlny. Četnost a ostrost „zubů“ v rámci jedné periody poukazuje na silnou přítomnost vyšších harmonických frekvencí.



Obr. 3.1. 25 ms zvukového signálu ženy (vlevo) a muže (vpravo). Převzato z Gussenhoven (2004: 2).

Z oscilogramu si lze udělat hrubou představu o výšce hlasu, ale o kvalitě lépe vypovídá, když si energii složek hlasivkového signálu zobrazíme ve spektru. Zde na vodorovnou osu vynášíme frekvenci v hertzích (Hz) a na svislou osu energii vyjádřenou v decibelech (dB), tedy jednotce, která nejlépe vystihuje relativní změny v amplitudě signálu vůči určité referenční hodnotě.

Jelikož spektrum rozkládá zvukový signál na jednotlivé frekvenční složky (stejně jako lidské ucho), odpovídá lépe tomu, jak zvuk vnímáme. Najdeme v něm jak základní frekvenci, tak i její násobky – vyšší harmonické frekvence, jejichž rozestup se od F0 odvíjí (u vyšších hlasů bude větší, u nižších menší). Jednotlivé složky o různých amplitudách dávají dohromady tvar spektra, který chápeme jako spektrální obálku. Ta se s ubývající energií ve vyšších frekvencích svažuje. Jedná se o důležitou vlastnost spektra, která se nazývá *spektrální sklon* a je korelátem vnímané barvy a kvality hlasu.

Pokud bychom si zobrazili spektrální složení zvuku, který vychází přímo z hlasivek, byl by spektrální sklon strmý (literatura uvádí -12 dB/oktávu), což napovídá tomu, že se v nízkých frekvencích energie soustředí hodně, a málo ve vysokých. Navíc se dá předpokládat, že syrový zvuk vycházející z hlasivek je velmi silný, např. Neppert (1999: 127) odhaduje 130-140 dB. Obě tyto skutečnosti jsou v rozporu se zastaralou fonetickou představou, že zvuk, který hlasivky vydávají, je slabý a řezavý.

Výsledný lidský hlas však dostává svou charakteristickou kvalitu, až když hlasivkový signál projde měkkým a vlhkým prostředím vokálního traktu. Ten má zhruba tvar polouzavřeného tubusu, a čím je delší, tím níže položená bude mít rezonanční pásma, tedy zesílené frekvence.

Vlivy na výsledný zvuk vycházející z retní štěrbiny popsal už v roce 1960 Gunnar Fant ve své *filtrové teorii produkce řeči*. Podle této teorie působí vokální trakt spolu s retní štěrbinou na hlasivkový tón jako filtr, který mu ubírá na hlasitosti a mírně vyrovnává spektrální sklon na výsledných -6 dB/oktávu.

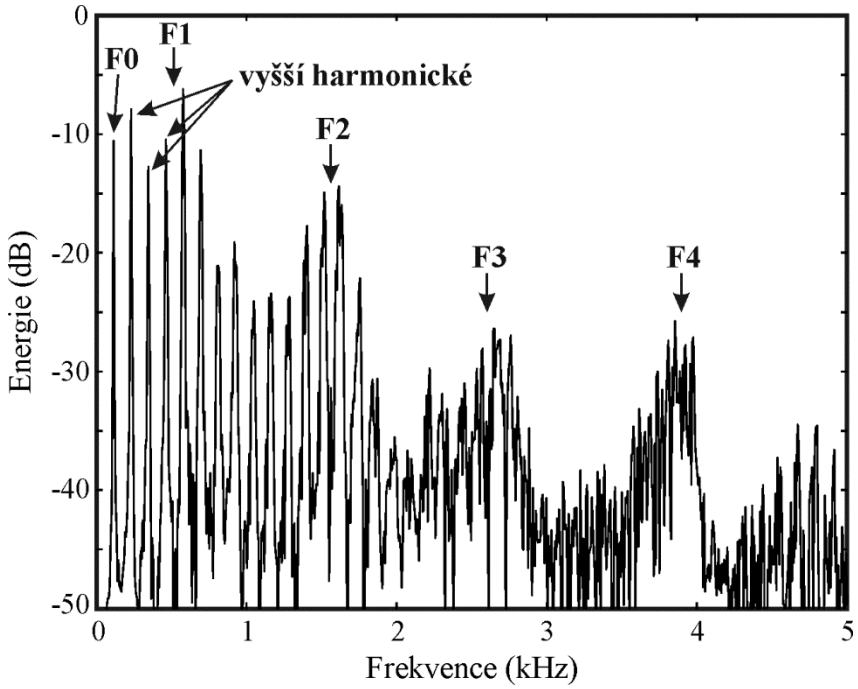
Do jisté míry barvu hlasu nelze ovlivnit, protože nadhrtanové dutiny modifikující hlasivkový signál – zejména nosní a přilehlé – mají svůj tvar daný, ale v ústní dutině lze velmi malými rozdíly v nastavení čelisti a jazyka docílit změn v rezonančních frekvencích a měnit tak samohlásky (nejlépe si to člověk uvědomí při artikulaci samohláskové řady, [aeɪou]). Vokalické formanty, tedy rezonance traktu charakteristické pro jednotlivé samohlásky, jsou

nejvýznamnějším spektrálním projevem filtrace hlasivkového tónu vokálním traktem. Ve spektrální obálce se projevují jako „kopečky“ zesílených harmonických složek signálu (viz obrázek 3.2). Je dobré si připomenout, že přítomnost vokalických formantů neznamená, že se v těchto frekvenčních pásmech zvuk ve vokálním traktu zesiluje, nýbrž jen o něco méně tlumí.

Jak jsme uvedli, vokalické formanty mají své typické konfigurace pro samohlásky. Kromě kvality dané samohlásky (tj. zda se jedná např. o přední zavřený vokál /ɪ/ nebo střední otevřený vokál /a/) závisí formantové hodnoty také na pohlaví nebo věku mluvčího. Například neutrální vokál [ə] (šva) u dospělého muže charakterizuje první formant na 500 Hz, druhý na 1500 Hz a třetí na 2500 Hz. U žen jsou tyto hodnoty však násobené koeficientem 1,2 (tedy 600 Hz, 1800 Hz a 3000 Hz), u dětí dokonce 1,6 (tedy 800 Hz, 2400 Hz a 4000 Hz). Vokalické formanty jsou ve fonetické literatuře hojně zkoumány, obzvláště první tři, protože v jejich dlouhodobých hodnotách pro individuální vokály i v jejich průbězích v reálném čase lze najít známky idiosynkratického řečového chování.

Vrátíme-li se k barvě lidského hlasu, bylo dokázáno, že měkčí/dyšnější hlas koreluje s nižší základní frekvencí (Laver, 1980). Proč tomu tak je, není přesně známo, ale nejspíš se udržení fonace za vyšší přítomnosti šumu lépe udržuje u nižších frekvencí.

Rozdíl mezi ostřejším a měkčím hlasem lze rovněž dobře pozorovat ve spektru – konkrétně ve spektrálním sklonu, a to tak, že ostřejší hlas má spektrální sklon méně strmý. Britta Hammarbergová a kolektiv autorů dokonce prokázali, že u dyšnějších hlasů dochází k propadům spektrální obálky přímo v oblastech mezi 0-2 kHz a 2-5 kHz (Hammarberg et al., 1980; citováno ve Weingartová, Bořil, Vaňková, 2014: 85). Jak uzavírají tito autoři svou kapitolu, spektrální analýzy a konkrétně spektrální sklon nabízejí výraznou pomoc při fonetické identifikaci mluvčího, ale je nutné mít dostatečné a srovnatelné zastoupení vokálů (tedy nejvíce znělých úseků signálu) v nahrávkách a zároveň musejí být nahrávky velmi kvalitní.



Obr. 3.2. Spektrum znázorňující spektrální čáry základní frekvence F0 a vyšších harmonických. Vrcholky vyšších hodnot amplitud odpovídají formantům F1, F2, F3 a F4. Převzato z Weingartová, Bořil, Vaňková (2014: 80).

3.2 Výška hlasu a inherentní F0 segmentů

Tak jako je spektrum signálu korelátem vnímané kvality a barvy hlasu, je základní frekvence akustickým korelátem vnímané výšky hlasu. Stejně tak platí naopak, že výška hlasu je percepčním korelátem F0. Nyní se podíváme blíže na to, jak se realita objektivní měřitelné F0 odráží v subjektivní doméně percepce hlasu.

Hlasivkový tón je primárně přítomný při artikulaci vokálů a znělých konsonantů. Z toho vyplývá, že při artikulaci neznělých konsonantů (např. explozív /p, t, k/) dochází k přerušení kmitání a tedy i intonační křivky. Lidský mozek však dokáže tato přerušení v řádu desítek milisekund automaticky doplnit pomocí extrapolace pohybu F0. K tomu, aby si lidské ucho přerušení skutečně povšimlo, musí být delší než 200 ms (Nooteboom, 1997). Stejně tak tento

nevědomý proces propojování intonační křivky naruší, když je rozdíl mezi F0 před a po přerušení příliš markantní.[1]

Na segmentální úrovni řeči dochází k jevům a procesům, které si při percepci ani neuvědomujeme a které neovlivňují obsah přenesené informace, protože nemají funkční platnost. Těmto jevům říkáme mikroprosodické. Představíme si je podle segmentů, ve kterých k nim dochází.

V rámci vokálů můžeme pozorovat inherentní základní frekvenci, která souvisí s různou vokalickou kvalitou. Vysoké vokály /u/ a /ɪ/ mají v rámci mluvčího konzistentně F0 vyšší o 4 – 25 Hz než nízký vokál /a/. Tento rozdíl je obzvláště výrazný v přízvučných slabikách. Proč k tomu dochází, vysvětluje více teorií. John Ohala a Brian Eukel (1987) provedli experiment, na němž ukázali, že vyšší napětí jazyka při výslovnosti vysokých vokálů zvedá hrtan a má vliv na hlasivkové napětí (tzv. *tongue-pull hypothesis*). Tento drobný rozdíl dokáže mozek vyfiltrovat, a naopak pokud má /a/ stejnou F0 jako /ɪ/, vnímáme /a/ jako vyšší.

Jak jsme popsali výše, u neznělých konsonantů je glotis otevřená, a tudíž není přítomná žádná F0. I u znělých konsonantů však může docházet k jejímu vymizení, protože při artikulaci dochází k uzavření nebo zúžení vokálního traktu, což způsobuje zvýšení tlaku vzduchu v ústní dutině a snížení transglotálního tlaku potřebného k udržení fonace. Tím se sníží frekvence vibrace, ale pro lidské ucho to není slyšitelné.

Rovněž bylo dokázáno, že ve vokálu následujícím po neznělém konsonantu začíná F0 o něco výše než po znělém. Jak píše Gussenhoven (2004: 7), je to tím, že otevření glotis v přípravě na neznělý konsonant vyžaduje vyšší svalové napětí v hrtanu, což se odrazí v bezprostředně následujícím segmentu vyšší frekvencí.

[1] Přerušení průběhu základního tónu není jediný případ, ve kterém lidský mozek dokáže doplnit chybějící informaci. Děje se tak i tehdy, když v signálu F0 úplně chybí. To nejběžněji nastává při telefonickém hovoru přes pevnou telefonní linku, protože přenos signálu je umožněn v pásmu 300 a 3400 Hz. Spodní frekvenční hranice vylučuje mužské i většinu ženských základních frekvencí, a tak se ucho posluchače musí spolehnout na rozestup harmonických frekvencí, z nichž si F0 odvodí, protože každá harmonická je násobkem F0.

4 Percepční aspekty základní frekvence

4.1 F0 v průběhu nádechového úseku

Výše popsané drobné změny v F0 mají svá fyziologická vysvětlení. Podobné je to u jiného jevu, který Jacqueline Vaissièrová považuje za univerzální vlastnost lidské řeči, společnou všem jazykům: deklinace F0. V průběhu promluvového úseku (proneseného běžným oznamovacím způsobem), který se kryje s nádechovým úsekem, křivka F0 neustále klesá a dosahuje čím dál nižších maxim i minim (viz Vaissière, 2005: 247). Způsobuje to úbytek vzduchu v plicích a tím i subglotálního tlaku, což s projevuje deklinací F0, i přes přirozené mechanismy svalové kompenzace. Po nádechu se výška F0 opět resetuje.

Všechny tyto faktory ovlivňující tvar intonační křivky jsou však automaticky percepcí filtrovány a při sestavování modelových intonačních kontur bývají vypouštěny jako uživatelsky zanedbatelné. Řeč je např. o nizozemském modelu IPO (viz např. Nooteboom, 1997; ´t Hart et al., 1990) založeném na *close-copy stylizations*, což jsou synteticky vytvořené intonační kontury se dvěma základními požadavky: percepční ekvivalence ověřená v percepčním testu a co nejmenší množství pohybů F0, které ještě k této ekvivalenci postačují. Cílem IPO modelu byla postupná abstrakce až na základní intonační vzorce, které tvoří kategorie, do nichž posluchači rozdělují povrchové realizace frází. Pro příklad uveďme, že angličtina nebo holandština mají těchto základních intonačních vzorců 6, ruština až 10 (Nooteboom, 1997: 7).

4.2 Intonační funkce F0

Máme-li definovat pojem intonace, narážíme na problém, s nímž se často fonetická literatura potýká. Intonace se dá v širším smyslu pojmout jako synonymum k prozodii, tedy suprasegmentální modifikaci řeči zahrnující kromě modulace F0 také modulaci silovou (kombinací obojího se dosahuje percepčního dojmu přízvuku v češtině) nebo temporální (prodlužování koncových slabik, pauzy). V užším smyslu se intonace chápe pouze jako relativní změny F0, které jsou mluvčím zamýšlené, posluchačem vnímatelné (fungující na bázi kontrastů) a lingvisticky významné, tedy funkčně zatížené. Pojmy „intonace“ a „melodie“ by se však neměly používat zaměnitelně.

Výčet nejdůležitějších funkcí intonace obsahuje umožnění posluchači orientaci ve výpovědi, tedy naznačení hranice promluvových celků (frázování, gramatická funkce), umístění prominence na přízvučnou slabiku, zvýraznění nové informace – rématu oproti známé – tématu (diskurzní funkce) a dále signalizaci předání slova druhému účastníkovi komunikační situace. Intonace také charakterizuje komunikační platnost promluvových celků (tedy zda se jedná o větu tázací či oznamovací) a v neposlední řadě obsahuje důležitou informaci o postojích a citových rozpoloženích mluvčího (Vaissière, 2005: 237 nebo také Roach, 2006: 183-184).

Již jsme naznačili, že hlasová a silová modulace umístěné na první slabiku slovního taktu se podílejí na percepci českého přízvuku. V jiných jazycích může být přízvuk realizován i temporálním kontrastem či může korelovat s určitou kvalitou vokálu v jádru slabiky. Pokud je v určitém jazyce kontrast realizován ustálenou modulací F0, jedná se o tónový přízvuk, který je prostředkem dosahování lexikálních rozdílů v tónových jazycích (např. čínština, vietnamština) nebo v jazycích s tónovým přízvukem, kde ustálený průběh F0 rozlišuje minimální páry v části slovní zásoby (např. norština, švédština, litevština).

4.3 Percepce F0 a nejvhodnější percepční škála

Na tomto místě je vhodné uvést ještě několik poznámek k posluchačově vnímání hlasu. Lidské ucho je schopné rozlišit poměrně drobné změny ve frekvenci tónu, ale tato schopnost je závislá na tom, v jak vysokých frekvencích se tento tón pohybuje. Do 800 Hz dokážeme rozlišit kroky tak jemné jako 3 Hz (*just noticeable difference,* tedy *sotva postřehnutelný rozdíl*). Od 800 Hz výš už je naše vnímání poměrové, což znamená, že rozdíl 2000 Hz ku 1000 Hz nám připadá stejný jako rozdíl 4000 Hz ku 2000 Hz. Nad 800 Hz už musí být krok ve frekvenci asi 0,5 %, abychom si ho všimli.[2]

Z toho vyplývá, že je potřeba o vnímání výšky hlasu uvažovat relativně, nikoli absolutně, a k tomu se frekvenční jednotka Hertz příliš nehodí. Proto byly zavedeny stupnice, které lidské percepci lépe odpovídají, protože stejně velký interval v rámci jedné škály odpovídá stejné

[2] Vnímání výšky hlasu není závislé jedině na změnách v základní frekvenci, může záviset také na amplitudě (síle) hlasu. U hlubších frekvencí dochází při větší hlasitosti k dojmu prohloubení tónu, u vyšších frekvencí naopak k dojmu dalšího zvýšení.

percepční vzdálenosti. Jedna z těchto stupnic nich je hudební půltónová škála, další je ERB škála, pak Barky a mely.

Jeden půltón odpovídá asi 6% změně ve frekvenci. Základem pro výpočet půltónu je logaritmus poměru příslušných dvou frekvencí. Každou oktávu lze rozdělit na 12 půltónů, u kterých pak pracujeme s jistotou, že změna o jeden takový půltón výš nebo níž bude vnímána jako stejně velký krok. Literatura k tématu (Rietveld a Gussenhoven, 1985) doplňuje, že vzdálenost 1,5 půltónu už je dostatečně markantní a prozodicky významná na to, aby s ní mluvčí funkčně pracoval při intonační modulaci.

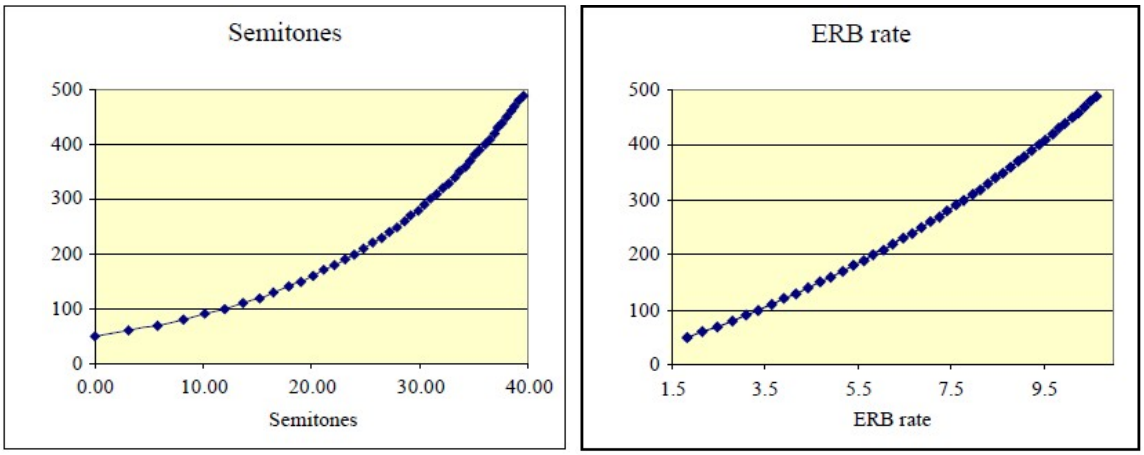
Stupnice založená na ERB (*Equivalent Rectangular Bandwidth*) jde ještě dál a výpočet jednoho kroku zakládá na znalosti vlastností bazilární membrány ukryté ve vnitřním uchu. Na té je každá frekvence detekována na určitém malém úseku se sluchovými buňkami. Frekvence, které jsou si příliš blízko, se mohou na jednom úseku slít dohromady, čímž nevytvoří percepční rozdíl. Pokud ale při souběžném poslechu frekvencí vzdálených o *jnd* detekujeme přítomnost cizí frekvence, spadá už do jiného úseky. Jeden takovýto úsek nazýváme *kritické pásmo slyšení* a ERB stupnice dokáže jeho šířku stanovit jako šířku pásma obdélníkového filtru, který propouští takové množství šumu, jaké ještě neinterferuje s daným tónem.

Bark je jednotka, která se určuje podobně jako jeden ERB, akorát „z druhé strany" – využívá se signálního tónu a úzkopásmového šumu kolem jeho frekvence. K maskování signálního tónu dochází, když je šířka šumového pásma rovná šířce sluchového filtru. Nevýhoda Barků, na rozdíl od ERBů, spočívá v hrubém rozlišení v nízkých frekvencích.

Melová stupnice byla ve 30. letech minulého století odvozena za pomoci percepčních testů, u nichž byli účastníci instruováni, aby upravili výšku daného tónu na takovou, která podle nich odpovídá poloviční nebo dvojnásobné výšce. Takto se dospělo až ke škále rozsahu 0 – 2400 melů.

Podle Sieba Nootebooma (1999: 4), který se zabýval formanty ve spektru, stupnice ERBů lépe mapuje citlivost lidského ucha na prominence dosahované rozdílně u hlasových rejstříků mužů a žen. Experiment Francise Nolana (2003) zase ukazuje, že pro popis relativních změn výšky hlasu se lépe hodí půltóny. Ve své studii porovnal všechny nejpoužívanější systémy pro škálování výšky hlasu, a to tak, že nechal respondenty (muže i ženy) opakovat fráze tak, aby

bylo dosaženo co nejlepší intonační ekvivalence. Porovnání rozdílů mezi šablonou a opakováním, prováděné v Hertzech a čtyřech psychoakustických škálách, ukázalo, že přepočítáno na půltóny dosáhli respondenti nejlepších výsledků. To dokazuje, že půltónová škála nejlépe odpovídá intuitivnímu chápání ekvivalence mezi intonačními vzorci vnímaného a reprodukovaného. Pro lepší představu viz obrázek 4.1 níže.



Obr. 4.1. Vztah jednotky Hertz od 50 – 500 k příslušným půltónům (vlevo) a ERBům (vpravo). Převzato z Nolan (2003).

5 Forenzní analýza

5.1 Posuzování nahrávek

Forenzní fonetika se nejčastěji zabývá rozpoznáváním mluvčího na základě zvukových nahrávek. Toto rozpoznávání může být dvojí, a sice za účelem verifikace – ověření totožnosti mluvčího, které se však převážně odehrává automaticky a není při něm potřeba zásah člověka. Druhým účelem rozpoznávání je identifikace – určení totožnosti mluvčího, které se ve forenzní praxi dále dělí podle okolností konkrétního případu, tedy toho, zda je k dispozici nahrávka trestného činu (telefonátu o uložení bomby, zneužití nouzového volání nebo např. vydírání – viz Skarnitzl, 2014: 14) či nikoli. Pokud nahrávka existuje, ale zatím neexistuje podezřelý, slouží informace z této nahrávky k zúžení populační skupiny, z níž podezřelý pravděpodobně pochází. Provádí se tzv. profilování mluvčího. Hledají se především idiosynkratické zvláštnosti, jež mohou být organické povahy (vokální trakt), nebo osvojené (příslušnost k nářeční skupině, idiolektismy, zvláštní prozodické návyky, artikulační strategie, atp.).

Pokud podezřelá osoba (popřípadě více osob) existuje, přijde na řadu pořizování srovnávacích nahrávek a jejich analýza ve vztahu ke sporné nahrávce. Výstupem je pak vyjádření na základě pravděpodobnostní škály, s jakým stupněm jistoty se jedná či nejedná o shodnost mluvčího. Pravděpodobnostní škála může vypadat např. takto:

Je pravděpodobné na hranici jistoty

Je vysoce nepravděpodobné

Je pravděpodobné

Je nepravděpodobné

Je vysoce nepravděpodobné

Je nepravděpodobné na hranici jistoty

… že se jedná o tutéž osobu. (převzato z Bořil a Weingartová, 2014: 116)

5.2 Vliv telefonního přenosu na F0

Sporná nahrávka často nejdříve projde přenosovým kanálem, který signál zakóduje tak, že na výstupu již nemůže být rekonstruován ve vysoké kvalitě. Ačkoli se technologie neustále vyvíjejí a zdokonalují, stále může být řeč o nikoli nepodstatném vlivu telefonního přenosu na nahrávku.

Vaňková a Bořil (2014: 104) shrnují možné vlivy telefonního přenosu jako trojí: a) vlivy okolí, b) vlivy pocházející od mluvčího a c) technické vlivy. Mezi vlivy pocházející od mluvčího patří jak úmyslná modifikace hlasu – tzv. maskování – tak neúmyslná změna rejstříku, známá jako typický *telefonní hlas*, mezi jehož příznaky patří změna kvality hlasu, artikulačního tempa, nebo, jak si popíšeme dále, zvýšení průměrné F0.
Mezi technické vlivy, o kterých je především řeč, patří hlavně pásmové filtrování přenosu. Autoři zmiňují AMR kodeky, které se používají při digitalizaci přenášeného zvuku v mobilních telefonech. Na rozdíl od pevné telefonní linky úzkopásmový AMR kodek propouští větší rozsah frekvencí (100 Hz až 2800 – 3600 Hz), širokopásmový dokonce 50 Hz – 7000 Hz (tamtéž: 107). Autoři ale v oddíle věnovaném vlivu kodeků na F0 varují, že není možné predikovat, zda kodek pro daného mluvčího mění distribuci hodnot F0 (průměr a směrodatnou odchylku), či nikoli (tamtéž: 112).

Co se týká vlivu pevné telefonní linky, Hirson et al. (1995) v rámci své studie pozoroval její vliv na výsledné průměry F0. V závěru uvádí, že hodnoty extrahované z telefonních nahrávek jsou průměrně o něco vyšší než ty ze spontánních rozhovorů. Jedná se však o rozdíly v jednotkách Hz, proto tento závěr nemá pro forenzní fonetiku příliš velkou váhu. Co však stojí za zmínku je postřeh autorů, že sporné telefonní nahrávky, jež jsou předmětem fonetické analýzy, obvykle obsahují citový náboj, což střední hodnoty F0 také zvyšuje. Jestliže je tedy u podezřelého subjektu naměřena nižší F0 než v inkriminující telefonní nahrávce, existuje vysoká pravděpodobnost, že se nejedná o tutéž osobu (tamtéž: 237).

Tuto hypotézu chtěli ve svém výzkumu ověřit i Gfroerer a Wagner (1995), kteří měli od jednoho mluvčího k dispozici tři nahrávky reálných telefonických výhrůžek, jejichž průměrné F0 se pohybovaly kolem 170 Hz. Ve vazbě byly potom pořízeny další nahrávky s průměry nižšími, okolo 140 Hz. Tento významný 30Hz rozdíl dále ověřili na materiálu od 76 německých mluvčích (telefonní nahrávky kriminální povahy) porovnaném s daty od 100

německých mluvčích (z laboratorních nahrávek čtených výhrůžných textů, Künzel, 1987). Rozdíl průměrných F0 mezi těmito skupinami byl 27 Hz. Závěr, který z těchto poznatků autoři vyvodili, byl následující: Při forenzním srovnávání sporné a srovnávací nahrávky se musí pamatovat na významnou tendenci mluvčího mít vyšší hlas ve sporné reálné nahrávce, a to buď kvůli snaze přehlušit okolní šum, kvůli snaze dát najevo sebejistotu a dominanci (srov. vyšší F0 v hovoru nadřízeného směrem k podřízenému, Zraick *et al.*, 2006) nebo kvůli stresu vyplývajícímu ze situace.

5.3 Sluchově-percepční analýza

Základní způsoby, kterými se nahrávky analyzují, jsou dva: poslechová a akustická analýza. Jak píše Skarnitzl (2014: 16), tyto dva postupy jdou ve forenzní fonetice ruku v ruce, protože poslechová analýza prováděná fonetickým odborníkem dokáže v nahrávce odhalit jevy, které akustická analýza nezachytí – a naopak. Dialekt, cizinecký přízvuk a zvláštnosti osobního jazykového systému jsme mnohem lépe schopni zhodnotit poslechem, kdežto akustické parametry (F0, formanty a jejich statistiky) musí extrahovat počítač. Na akustickou analýzu se blíže podíváme v kapitole 7.

Co se týče sluchově-percepční analýzy, nejpropracovanější systém roztříděných charakteristik vyvinul Hollien (2000). Jedná se o sadu 7 parametrů (výška hlasu, kvalita hlasu, intenzita, dialekt, artikulace, prozodie a ostatní – dysfluence a řečové vady), z nichž každý má ještě 1 – 5 dílčích charakteristik (např. artikulace: vokály, konsonanty, nazalita, chybná artikulace). Při porovnávání dvou nahrávek postupuje fonetik po jednotlivých charakteristikách, při poslechu se soustředí jen na ně a na stupnici 0 (nejmenší podobnost) – 10 (největší podobnost) dané dvě nahrávky srovnává. Označená skóre se na konci analýzy sečtou a vyjádří se jako procento z celkového počtu bodů, kterého by se dosáhlo při maximální podobnosti.

Podle Mackenzie Beck (2005) je nejužitečnější užívat podobné hlasové profily (vytvářené za účelem co nejlepšího postihnutí komponent kvality hlasu) v oblasti hlasové patologie, protože napomáhají stanovit typické symptomy té které hlasové poruchy a také sledovat u pacienta účinek používané terapie. Zároveň argumentuje, že forenzní použití takovýchto hlasových profilů může být v některých případech ošemetné. Provedla 120 párových srovnání

profilů od 50 skotských mluvčích a zjistila, že 14,2 % případů by se dalo označit za hlasové dvojníky. Z toho usoudila, že by se jen na shodnosti dvou hlasových profilů nemělo zakládat tvrzení o shodě mluvčích. Domníváme se však, že tento závěr není příliš relevantní, protože ve forenzní praxi nemíváme tolik srovnávacích nahrávek, a tedy pravděpodobnost výskytu hlasových dvojníků je nižší.

5.4 Likelihood ratio

Jakmile jsou na nahrávkách provedeny potřebné analýzy (poslechové i akustické – viz kapitola 7), je možné s nimi pracovat jako s důkazním materiálem a interpretovat jejich váhu tak, aby ji šlo formulovat během soudního řízení. K tomu může posloužit *likelihood ratio,* neboli česky věrohodnostní poměr. Tento koncept pracuje s tzv. bayesovskou statistikou, která umožňuje se vyjádřit k výrokům a hypotézám, jejichž pravdivost je neznámá, a o jevech, které se neopakují (na rozdíl od tzv. frekventistické statistiky, která zpracovává velké objemy dat). To je vhodné v případě, kdy se forenzní fonetik chce u soudu vyjádřit k důkaznímu materiálu a k jeho kompatibilitě s hypotézou obžaloby H_O (že nahrávky pachatele a podezřelého pocházejí od jedné osoby) a s hypotézou obhajoby neboli alternativní hypotézou H_A (že nahrávky pocházejí od různých osob).

$$\frac{p(H_O|D)}{p(H_A|D)} = \frac{p(H_O)}{p(H_A)} \times \frac{p(D|H_O)}{p(D|H_A)}$$

aposteriorní pravděp. apriorní pravděp. věrohodnostní poměr

Obr. 5.1. Vzorec Bayesovy věty, převzato z Bořil a Weingartová, 2014: 128.

Věrohodnostní poměr je součástí výše uvedené rovnice, v níž se pracuje s apriorní pravděpodobností na straně jedné (formulované na základě dosavadního vyšetřování před tím, než se přistoupí k analýze důkazního materiálu, tedy řečových nahrávek) a s aposteriorní pravděpodobností na straně druhé, což je apriorní pravděpodobnost přenásobená právě věrohodnostním poměrem po zvážení důkazu. Věrohodnostní poměr vyjde jako číslo, o němž se dá vyjádřit například následovně: „*Když vezmeme v úvahu rozdíl mezi vzorky řeči*

podezřelého a pachatele, je tisíckrát pravděpodobnější, že takový rozdíl vznikne, pokud budou oba vzorky pocházet od stejného mluvčího, než pokud pocházejí od různých mluvčích" (podle Rose, 2006: 162; citováno v Bořil a Weingartová, 2014: 130).

Pokud by náhodou apriorní pravděpodobnost byla 0, protože podezřelý mluvčí má na daný trestný čin alibi, i po vynásobení tisícem by aposteriorní pravděpodobnost byla stále 0. Věrohodnostní poměr < 1 podporuje hypotézu obhajoby (nejsou jedna osoba) a pokud je rovný 1, nevyjadřuje se ani k jedné hypotéze.

Apriorní pravděpodobnost pro daný případ by forenzní expert neměl znát, aby nebyl předem ovlivněný a vyjadřoval se výhradně k důkazu. Měl by si také dát pozor, aby se vyjádřil jen k pravděpodobnosti důkazu, nikoli k pravděpodobnosti hypotézy – to může na základě aposteriorní pravděpodobnosti udělat soud. Mezinárodní asociace pro forenzní fonetiku a akustiku (IAFPA) v roce 2004 zveřejnila etický kodex pro své členy, který upravuje mimo jiné i to, že musejí vždy objasnit metody své analýzy a vysvětlit jejich opodstatněnost (http://www.iafpa.net/code.htm).

5.5 Diskriminační index d‘

Výše popsaná vyjádření ke shodnosti/neshodnosti mluvčích v nahrávkách jsou obvykle prací fonetiků-odborníků. Vezmeme-li v úvahu, že v praxi je možné použít i výpověď svědků, kteří v dané situaci slyšeli pachatelův hlas, přichází na řadu otázka, jak dalece je taková výpověď spolehlivá.

Bořil a Weingartová (2014: 118) citují autory Schillera, Köstera a Duckwortha (1997), kteří testovali sluchovou paměť respondentů tak, že seznámili své respondenty s hlasem cílového mluvčího a poté jim pouštěli vzorky řečí (bez lingvistické informace). Použitým testem byl 2AFC test (*Two-alternative forced-choice test* neboli *nucená volba ze dvou alternativ;* viz např. tamtéž: 118), v němž se uvádí rozhodnutí ANO či NE na otázku, zda se u dvou nahrávek jedná o téhož mluvčího. Tato rozhodnutí spadají vždy do jedné ze čtyř kategorií:

Stimul ↓ Odpověď →	**Ano**	**Ne**
Cílový mluvčí	správně pozitivní	chybně negativní *(miss)*
Jiný mluvčí	chybně pozitivní	správně negativní *(correct rejection)*

Tabulka 5.1. Typy odpovědí u 2AFC testu, v němž posluchači odpovídají „ano“, pokud slyší shodný hlas, a „ne“, pokud jiný. Převzato z Bořil, Weingartová, 2014: 118.

Odpovědi jednotlivých respondentů lze poté vyhodnotit pomocí tzv. *diskriminačního indexu d‘*. Ten dobře popisuje, jak je respondent citlivý na hlas cílového mluvčího, a jestli je v jeho mysli jednoznačně oddělený od ostatních necílových vzorků. Počítá se za použití míry úspěšnosti H (podíl správných odpovědí ze všech odpovědí, *hits*) a míry falešných poplachů (*false alarms*) následujícím způsobem:

$$d' = Z(H) - Z(F)$$

Funkce Z neboli z-skóre zde počítá vzdálenost od průměru měřenou ve směrodatných odchylkách. Čím je výsledný diskriminační index d‘ vyšší, tím úspěšnější je respondent v rozlišování mezi hlasy. V ideálním případě by diskriminační index d‘ byl roven 1.

6 Střední hlasová frekvence a faktory, které na ni mají vliv

6.1 Střední hlasová frekvence (SFF)

Každý člověk má určitou obvyklou hlasovou polohu nebo také střední hlasovou frekvenci (*speaking fundamental frequency*), kolem níž se většinu doby jeho hlas pohybuje. Jak píší Skarnitzl a Hývlová (2014), z literatury není zcela jednoznačné, jak se SFF počítá, ale v největším počtu případů zřejmě jako průměrná F0 z delšího úseku řeči, dále se může počítat jako medián (prostřední hodnota na škále vzestupně nebo sestupně seřazených hodnot) či modus (nejčastěji se vyskytující hodnota) (tamtéž: 49). Je také spojována s určitou směrodatnou odchylkou, která značí hlasovou variabilitu.

Velikost této směrodatné odchylky samozřejmě může ovlivňovat řada faktorů od momentálního citového rozpoložení pro dlouhodobé návyky (řečové, e i neřečové, např. kouření). Některé národy mají směrodatnou odchylku SFF obecně větší, např. právě tónové jazyky, jako je čínština. Ke zvyšování směrodatné odchylky, tedy variabilnější práci s hlasem, dochází podle některých studií také s věkem (Pegogaro-Krook, 1988, citováno v Traunmüller a Eriksson, 1994).

6.2 Změny SFF v závislosti na věku, hluku, denní době a psychickém rozpoložení

Samotná průměrná hodnota SFF se také s věkem mění, ačkoli badatelé se rozchází v tom, zda více u mužů, či u žen, a také jakým směrem; podle Hollienovy studie (1987, citováno v Nishio a Niimi, 2008) dochází k větším změnám u mužů, a sice ke zvyšování SFF, kdežto autoři Nishio a Niimi na svém materiálu zjistili větší změny u žen, a to směrem dolů – a to přibližně o 10 Hz každé desetiletí, rozdíl mezi průměry skupiny 20letých žen a 80letých žen pak byly 226 Hz a 168 Hz (tamtéž: 123). Dále se SFF zvyšuje při vyšším mluvním úsilí, jak zkoumal např. Jessen *et al*. (2005) použitím Lombardova jevu, při němž respondent ve sluchátkách slyší rušivý šum o hlasitosti 80 dB, což ho nutí instinktivně zesílit vlastní hlas, aby šum přehlušil. V této studii se jednalo o průměrné zvýšení ze 120 Hz za normálního mluvního úsilí na 159 Hz. Tytéž výsledky přinesla i studie autorů Bořil a Hansen (2011; citováno v Skarnitzl a Hývlová, 2014: 53). Dalším faktorem, který může za zvyšování průměrné F0, je

podle studie Garrett a Healey (1987; citováno tamtéž) denní doba. V jejich výzkumu měli muži při odpoledním nahrávání konzistentně vyšší F0 než při ranním.

Ve fonetické literatuře už bylo rovněž hojně popsáno, jaké vlivy mají na F0 různé emoční stavy. Studie se obecně shodují na tom, že radost či vztek se projevují vyšší F0 i vyšší variabilitou, naopak u smutku je obojí nižší. Obecně bývá F0 vyšší také u stresu, ale není to zákonitá tendence, mluvčí se v tom mohou lišit. Braun (1995) v sekci věnované psychologickým faktorům majícím vliv na F0 odkazuje na studie, které tento rozporuplný vliv stresu na F0 potvrzují, např. od kolektivu autorů Hecker *et al*. (1968), kteří jako spolehlivější ukazatel stresu navrhují třas hlasu (*jitter*). Pro forenzní praxi je každopádně dobré, když je možné porovnávat nahrávky, které byly pořízeny za co nejpodobnějších podmínek, a to i afektivních; u srovnávací nahrávky to obnáší simulaci stresové situace (Boss, 1996; citováno v Skarnitzl a Hývlová: 52).

Dosažení takové podobnosti může samozřejmě překážet, pokud je ve sporné nahrávce použito úmyslné maskování hlasu. To může být buď elektronické, za použití hlasových modulátorů, nebo neelektronické, nejčastěji pomocí zacpání nosu, předmětu v ústech, šepotu, sípání, chrapotu, fistule, prohloubení hlasu, simulace vady artikulace některých hlásek, předstírání akcentu jiného jazyka, nebo velmi pomalého tempa řeči často kombinovaného s jinou deformací (Svobodová a Voříšek, 2014: 142).

6.3 Vliv různých typů úloh na SFF

V laboratorních podmínkách záleží také na typu diskurzu, který je v experimentech použit. Zraick et al. (2006) zaměřili svou studii na 6 různých řečových podmínek a na to, zda se v některé z nich SFF významně liší od jiných. Tyto podmínky zahrnovaly veřejný proslov, konverzaci se členem rodiny, konverzaci s vrstevníkem, hovor směrem k podřízenému a hovor směrem k nadřízenému. Poslední dvě komunikační situace vykazovaly největší odlišnost SFF, a sice vyšší SFF v prvním a nižší SFF v druhém případě. Samotní autoři však uznávají, že na jejich výsledky mohly mít velký vliv experimentální podmínky, neboť účastníci si své konverzační protějšky měli za úkol jen představit.

Vliv různých úloh zčásti zkoumal také Hollien (1997) a dospěl k závěru, že převládá tendence mít vyšší střední hlasovou polohu při čtení psaného textu než při spontánní mluvě. Dodává však, že na základě těchto výsledků nelze spolehlivě předvídat řečové chování mluvčího.

Skarnitzl a Vaňková (IAFPA, 2015) pozorovali 26 mluvčích – mužů – u 3 typů úloh: spontánní dialog (se skutečným protějškem), čtený text a text obsahující stejná spojení, předvedený s úmyslným maskováním hlasu. Všechny tři typy úloh přinesly zajímavé výsledky. Sice potvrdily Hollienův (1997) závěr, že u čteného textu je obvyklé, když má subjekt vyšší hlasovou polohu, ale oproti očekávání se ukázalo, že česká intonace je u spontánního řečového projevu poměrně plochá, a to jak na úrovni mluvního taktu, tak na úrovni celé fráze. Rozptyl hodnot byl u spontánní řeči o poznání menší než u čteného textu. Může za tím být příběhová povaha čteného textu, ale důležitý poznatek je ten, že bohužel spontánně mluvená čeština nevykazuje velké známky variability, které by mohly být využitelné ve forenzní praxi.

Vliv dramatičnosti čteného textu na hodnoty F0 potvrdila také studie již zmíněného Allena Hirsona a jeho kolegů (1995). Ti se rozhodli z nahrávek čteného textu vyeditovat pasáže s přímou řečí, a výsledné průměry F0 poté odpovídaly hodnotám získaným z nahrávek spontánních rozhovorů.

6.4 Populační průměr

Dalším zajímavým výstupem experimentu Skarnitzla a Vaňkové (2015) bylo, že čeští mluvčí ve věku 20-45 let mají průměrnou F0 132 Hz. Jakmile se takové zjištění provede u většího počtu osob, vytváří se tím u daného parametru tzv. populační statistika, která říká, které hodnoty tohoto parametru jsou u většiny populace obvyklé a které vzácné. Pro české muže v tomto věkovém rozmezí je populační průměr nezvykle vysoký, srovnáme-li ho s výše zmiňovanou studií Nisho a Niimi (2008), kde byla pro japonské muže v této věkové skupině 122 Hz, nebo v porovnání s německým průměrem 120 Hz, který byl však získaný pro skupinu mužů od 21 – 63 let (Jessen *et al.*, 2005). Hudson *et al.* (2007) provedli ve Velké Británii výzkum na homogenní skupině 100 mužů ve věku 18 – 25 let, v němž simulovali podmínky policejního výslechu. U této skupiny následně určili průměrný průměr, modus a medián F0 – 106 Hz, 102,2 Hz a 105 Hz. Zároveň se hodnoty průměrné F0 u 60 % mluvčích nacházely

v poměrně úzkém intervalu 99 – 120 Hz, což značí, že F0 pro tyto a podobné případy nemůže být spolehlivým forenzním prediktorem. Naproti tomu však účinně od většiny populace odliší ty případy, které do tohoto pásma nespadají (tamtéž: 1810).

6.5 Forenzní relevance SFF

Z forenzního hlediska nás ovšem zajímá, jestli má střední hlasová poloha a její směrodatná odchylka nějaký forenzně využitelný potenciál. Z tohoto pohledu se zkoumá variabilita mezi mluvčími (*between-speaker variability*) versus variabilita v rámci jednoho mluvčího (*intra-speaker variability*), s tím, že je záhodno, aby poměr obou byl co největší. Protože však velikost směrodatné odchylky podléhá tolika situačním faktorům, jak jsme popsali výše, zaměřují se forenzní fonetici častěji na střední hlasovou polohu (SFF).

Aby bylo možné porovnávat mezi sebou ženské a mužské hlasy, používají se k definici intervalů směrodatné odchylky nebo intervalů průměru často právě půltóny popsané v oddílu 4.3, případně se variabilita vyjadřuje variačním koeficientem (směrodatná odchylka/průměr*100, v %). Coleman a Markham (1991) se ve svém výzkumu snažili přijít na to, jaké rozpětí intervalu pro SFF ještě lze považovat za normální pro jediného mluvčího, čímž chtěli vyvrátit tvrzení některých starších studií (např. Brown et al., 1989) o statistické významnosti rozdílu SFF mezi dvěma skupinami, který činil 1,58 půltónu (přičemž směrodatná odchylka v rámci skupin byla 2 a 2,44 půltónu). Výsledkem jejich bádání bylo, že průměrná základní frekvence se v čase mění i v rámci jednoho mluvčího, tedy že průměr F0 pro jednoho mluvčího leží v určitém intervalu. Tento interval pomocí analýzy srovnatelných nahrávek pořízených s určitým odstupem určili jako +/- 3 půltóny. Pokud se mluvčí od své SFF vzdálí mimo toto rozpětí, lze se odůvodněně domnívat, že za tím jsou fyzické či psychické zdravotní potíže (Coleman a Markham, 1991: 176).

Tyto výsledky, vypovídající o intraindividuální variabilitě, by se měly brát v úvahu, zejména pokud chceme střední hlasovou frekvenci použít např. za jeden z prediktorů v rámci lineární diskriminační analýzy (jak je popsáno v kapitole 7 o druzích akustických analýz), která spoléhá na dostatečný odstup mezi obvyklými hodnotami daného parametru, aby se významně projevily rozdíly mezi daty od různých mluvčích.

6.6 Baseline – základní hladina

Doposud jsme popisovali nejrůznější vlivy, které můžou mít okolnosti řečového projevu na zkoumané hodnoty F0. Logicky zde vyvstává potřeba najít nějaký řečový parametr, který bude vůči takovým vlivům odolnější. Traunmüller (1994) úspěšně navrhl ve své modulační teorii řeči existenci určité neutrální, nosné frekvence, kterou mluvčí moduluje v rámci dosahování prozodických i jiných cílů, ale k níž se vždy vrací. Tato nosná frekvence obsahuje základní údaje o mluvčím, jako je jeho věk nebo pohlaví (Traunmüller a Eriksson, 1994). Kromě toho zůstává u mluvčího stejná nehledě na míru živosti a emotivnosti jeho projevu, a dokonce se nemění ani s vyšším řečovým úsilím. Traunmüller a Eriksson (2006; citováno v Lindh a Eriksson, 2007) tuto nosnou frekvenci pojmenovali *základní hladina* (ZH, anglicky *baseline*). Autoři Lindh a Eriksson (2007) v úvodu své statě poznamenávají, že na předchozím výzkumu již bylo ukázáno, že se zvyšující se živostí projevu přirozeně vzrůstá nejen směrodatná odchylka, ale i průměrná F0.[3] To znamená, že exkurze F0 od střední hlasové frekvence se na frekvenční ose nezobrazují symetricky podle průměrné hodnoty, ale musí existovat jiná, neměnná hlasová frekvence, která je blízká nejnižší poloze, na níž je mluvčí schopen udržované fonace – tedy právě základní hladina. Ze vztahů průměrné F0, směrodatné odchylky a minimální F0 v dané promluvě Lindh a Eriksson odvodili následující vzorec pro výpočet základní hladiny (2026):

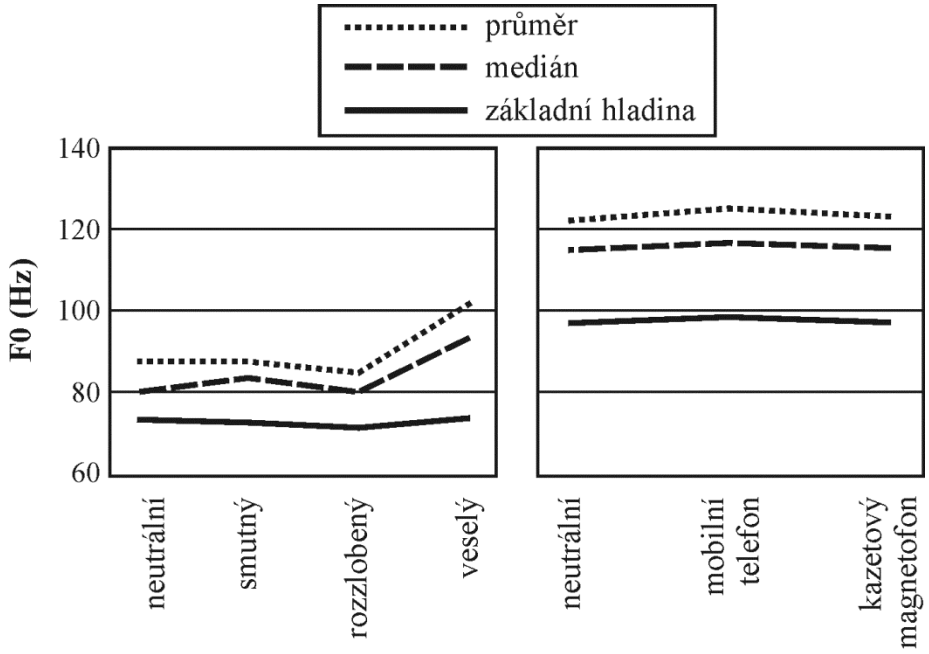
$$Fb = Fmean - k \cdot \sigma (F)$$

Dále ale upřesňují, že takový výpočet základní hladiny bude velmi citlivý na kvalitu nahrávky. V případě že v důsledku horší kvality nahrávky naroste počet chyb v extrakci F0, čímž se uměle zvýší směrodatná odchylka, je lepší zvolit postup výpočtu, který srozumitelně přeformulovali Skarnitzl a Hývlová (2014, 55): „Základní hladinu pro daný datový soubor stanovíme tak, že ve vzestupně seřazených hodnotách F0 identifikujeme frekvenci, která leží na percentilu 7,64; jinými slovy pod základní hladinou leží 7,64 % vzestupně seřazených hodnot."

[3] Tuto pozitivní korelaci popsal i Jessen *et al*., 2005, po srovnání průměrné F0 při neutrálním a vysokém mluvním úsilí, a zároveň uvedl přesvědčivý argument prosazující použití variačního koeficientu spíše než směrodatné odchylky vyjádřené v absolutních Hertzích. Variační koeficient (směrodatná odchylka/průměrná F0 * 100 [%]) se pak stává parametrem nezávisle ukazujícím skutečnou variabilitu, nehledě na změnu průměrné F0 vlivem jiného mluvního úsilí (194 – 195).

Lindh a Eriksson (2007) hypotézu o základní hladině podrobili třem druhům experimentu: 1. výpočet ZH u 5 různých simulovaných afektivních stavů, 2. výpočet ZH u identické nahrávky pocházející ze 4 různých přenosových kanálů nebo nahrávacích zařízení, 3. výpočet ZH pro 5 stupňů hlasového úsilí.

Jak je patrné z obrázku 6.1, ZH vykazuje stabilitu zejména v prvním a druhém experimentu. Ukazuje se jako parametr, který je jednak spolehlivější než tradičně počítané střední hodnoty, průměr a medián, jednak je odolný různé kvalitě nahrávacího zařízení.



Obr. 6.1. Porovnání základní hladiny, průměru a mediánu v experimentálních nastaveních autorů Lindh a Eriksson (2007). Vlevo: simulované afektivní stavy, vpravo: různé přenosové kanály. Upraveno podle Skarnitzl a Hývlová (2014: 55).

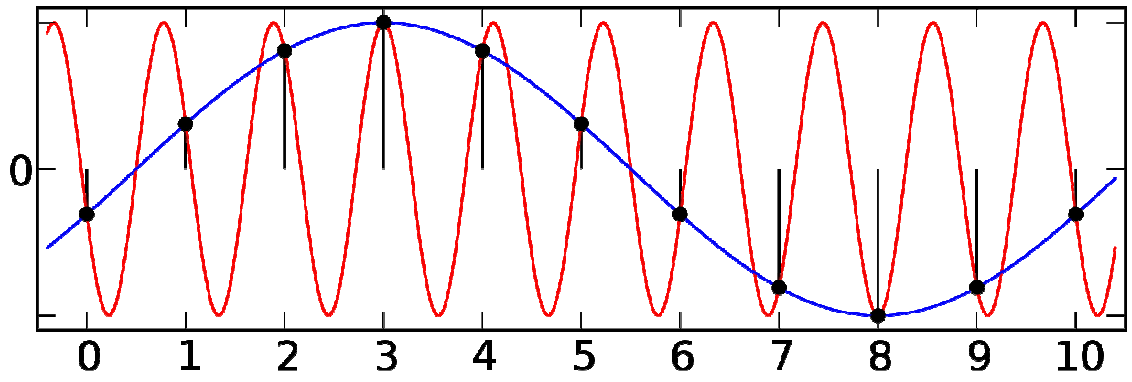
7 Druhy akustických analýz

7.1 Extrakce F0

Nyní se podíváme blíže na akustickou analýzu hlasu. Ponecháme stranou charakteristiky, jako jsou vokalické formanty, spektrální sklon nebo temporální vlastnosti řeči a zaměříme se na základní frekvenci hlasu, na způsoby její extrakce a současně na různé metody statistického zpracování dat F0.

Před tím, než můžou být extrahovány hodnoty F0, musí být provedena digitalizace spojitého signálu na diskrétní. Součástí digitalizace je vzorkování, tedy odečítání určitého počtu hodnot signálu za sekundu. Tento počet se vyjadřuje jako vzorkovací frekvence a obecně platí, že by měla být více než dvojnásobná k maximální frekvenci objevující se v signálu (vzorkovací teorém). Jelikož je lidské ucho schopno vnímat zvuky o frekvenci až 22 kHz, používá se často vzorkovací frekvence 44,100 Hz, zejména pro hudební záznamy, které by při nižší vzorkovací frekvenci byly ochuzeny. Úspornější vzorkovací frekvence 16 kHz se nicméně dá použít na řečový signál, protože nad hranicí 8000 Hz už se neodehrávají relevantní řečové jevy (Gussenhoven, 2004: 4).

Tím, že se dodrží vzorkovací teorém, nemůže docházet k tzv. *aliasingu*, tedy jevu, kdy je z nejvyšší harmonické frekvence obsažené v signálu odečítán vzorek dvakrát nebo méně než dvakrát během jedné periody, což vede k falešnému výpočtu této frekvence a jejímu zrcadlení do první poloviny frekvenčního rozsahu spektrogramu.



Obr. 7.1. Schematické znázornění aliasingu.
Zdroj: https://commons.wikimedia.org/wiki/File:AliasingSines.svg.

K extrakci hodnot F0 se používají nástroje nazvané *pitch trackers* (přestože, jak jsme si upřesnili v oddíle 3.2, výška hlasu a základní frekvence nejsou synonymické pojmy). Tyto nástroje jsou počítačovými algoritmy, které snímají hodnoty F0 v pravidelných intervalech. Hojně používaný *pitch tracker* je v programu Praat (Boersma a Weenink, 2014). Pro jeho správné fungování je nezbytné, aby právě byla správně nastavena vzorkovací frekvence a aby byla ze signálu odstraněna stejnosměrná složka, tedy aby byl signál vycentrován v oscilogramu kolem nuly. *Pitch tracker* potom využívá autokorelace, která hledá podobnosti signálu se sebou samým a jejímž cílem je určit, s jakou periodou se vzorky v signálu vyskytují na stejných místech ve svislém směru. Praat používá analyzační okénko, které ideálně obsáhne trojnásobek nejnižší frekvence, již chceme v signálu detekovat. Toto okénko je přenásobeno Hanningovým oknem, aby byly vyhlazeny intenzitní skoky na krajích okénka. Potom je okénko posouváno po časové ose, dokud algoritmus nenarazí na takový fázový posun, při kterém dosahují součty hodnot vzorků současného okénka a okénka v čase 0 maximálních hodnot (viz např. Boersma, 1993).

f0 = (1 / časový posun maxima) / vzorkovací frekvence[4]

Navzdory tomuto sofistikovanému mechanismu *pitch tracker* v Praatu není nástroj dokonalý. Může se v mapování F0 zmýlit například tak, že F0 umístí do šumového úseku neznělé frikativy, nebo naopak nedetekuje znělost tam, kde má. Dalším běžným problémem můžou být tzv. oktávové skoky – při třepené fonaci může nestandardní rozmístění vrcholků periody interpretovat jako dvojnásobnou frekvenci, nebo v jiných případech algoritmus ignoruje každý druhý vrcholek a dochází k půlení hodnot F0. (Gussenhoven 2004: 6). Tyto nepřesnosti musí manuálně zkontrolovat a opravit člověk.

Dalším způsobem, jakým se dá získat hodnota základní frekvence, je využitím kepstra. V tom případě je potřeba k výpočtu použít spektrální reprezentaci signálu. Ze samotného spektra F0 extrahovat nelze, protože by k tomu byl kvůli co nejlepšímu frekvenčnímu rozlišení nutný dlouhý úsek signálu. Ten by však nebyl z hlediska základní frekvence stacionární. Kepstrum tento problém řeší. Nelineární transformací na spektrum, zlogaritmováním a zpětnou transformací se dosáhne komplexního kepstra (přesmyčka slova

[4] Podle http://www.ucl.ac.uk/~ucjt465/tutorials/praatpitch.html; zpracováno podle Boersma, 1993.

spektrum), jež lze zobrazit v časové oblasti. Na levé straně takového zobrazení lze potom odečíst neperiodické složky signálu a na pravé straně periodické. Rychlá periodická změna průběhu spektra vlivem buzení se projeví jako špička v čase odpovídajícím základní hlasivkové periodě.

7.2 Statistické metody deskriptivní, inferenční a exploratorní

Jakmile se ze znělých úseků řečového signálu od jednoho mluvčího vyextrahují hodnoty F0, které lze přiřadit ke krokům v reálném čase (obvykle 5ms nebo 10ms kroky), lze je dále použít ke statistickému zpracování pomocí frekventistických metod.

Frekventistické metody ve statistice se dále dělí na 3 podskupiny. Jestliže chceme pouze postihnout a graficky znázornit ukazatele popisující chování dat, pak volíme metody **deskriptivní**, mezi něž řadíme střední hodnoty souborů dat nebo ukazatele variability. Pokud chceme z našich naměřených vzorků odhadovat obecnější chování zvolené proměnné a vlivy různých faktorů na ni, jedná se o metody **inferenční**. Pomocí nich se ověřují hypotézy o datech, totiž o normálním rozdělení dat či o shodnosti/různosti středních hodnot souborů, např. u souborů dat pocházejících od dvou mluvčích, tzv. t-testy. Dále mezi inferenční metody patří i ANOVA (*ANalysis Of VAriance*) – test používající analýzu rozptylů v datech, umožňující stanovit (především u vícefaktorových testů), jaké faktory mají na stejnost či různost dat největší vliv (např. faktor pohlaví, vzdělání, věku, ale i konkrétního mluvčího či kvalita jím artikulované hlásky, v rámci níž se proměnná zkoumá…). Významnost nalezených rozdílů mezi soubory dat se testuje tzv. post-hoc testy.

Zatřetí je možné použít metod **exploratorních**, které pokročilými technikami odhalují skryté informace v datech (Bořil a Weingartová, 2014: 122). Ve forenzní fonetice je hojně využívaná lineární diskriminační analýza (LDA), která má za úkol rozdělit vložená data do kategorií mluvčích na základě zadaných prediktorů (např. formantových hodnot snímaných v časových odstupech, přičemž každé časové místo platí za jeden prediktor) a výstupem je pak klasifikační úspěšnost daná v procentech. Jak píše Volín (2007b: 276), výpočetně má LDA mnoho společného s analýzou rozptylu. Váže se k ní však podmínka, že počet použitých vzorků musí představovat nejméně dvacetinásobek použitých prediktorů nebo kategorií, do nichž se třídí, proto se mnoho studií zaměřuje na to, jak dosáhnout co nejlepší klasifikační

úspěšnosti za použití co nejmenšího počtu prediktorů. Sahá se potom ke složitější parametrizaci naměřených dat do menšího počtu čísel popisujících chování vzorků. Ne všechny studie však vycházejí ve prospěch této metody, např. autoři Skarnitzl, Lazárková, Nechanský a Šturm (2014) uzavírají svou kapitolu konstatováním, že LDA založená na vokalických formantech za významné považuje i rozdíly mezi formanty uvnitř jednoho mluvčího, a toto nezachycení intraindividuální variability je podle autorů výraznou překážkou v použití LDA v praxi (tamtéž: 35).

7.3 Dynamické parametry F0

Jako lze u formantů zkoumat kromě dlouhodobých parametrů i dynamické průběhy formantových kontur, objevila se snaha o něco podobného i u základní frekvence. Myšlenka za touto snahou je taková, že opomenutím melodických průběhů se může ztratit určitá část idiosynkratické informace o mluvčím. Musí se však také vzít v úvahu fakt, že k extrakci dynamické informace o F0 je zapotřebí větší míry supervize a oprav ze strany experimentátora, protože musí více kontrolovat chyby extrakce a připravovat materiál ve vztahu k časové ose nebo lingvistickému obsahu dané promluvy. Pro forenzní praxi je tedy vhodné volit určitý kompromis mezi pořízením dostatečného množství dat a nutností tato data korigovat.

Autoři Volín a Bořil (2014: 65-76) se o takový kompromis pokusili ve třech experimentech: výpočet gradientu regresní přímky, koeficienty polynomické rovnice a funkční analýza hlavních komponent. V těchto experimentech převáděli kontury F0 různými způsoby na menší počet parametrů a zkoumali, jak mezi sebou korelují 2 promluvy od jednoho mluvčího s podobným lingvistickým obsahem (např. *Řekneš jim, co si myslíš* a *Řeknete jim, co si myslíte*). Tímto způsobem dosáhli toho, aby se případná idiosynkrasie projevila výhradně ve způsobu manipulace s F0. Všechny tři experimenty vedly k obdobnému výsledku, ačkoli jejich komputační náročnost je různá, a sice že vzájemná korelace parametrů korespondujících promluv je poměrně vysoká. U prvního a třetího experimentu se jednalo o hodnotu přes 0,7; u druhého experimentu dokonce 0,9, ale jen u koeficientu popisujícího začátek promluvy. Jak uzavírají samotní autoři, jejich experimenty narážejí na problém s interpretací, protože zatím není známo, které melodické úseky promluvy mluvčí vědomě plánuje a které postupy jsou

náhodné (tamtéž: 76). I když je tento směr práce s extrahovanými hodnotami F0 tomto směru bude tedy potřeba ještě nadále bádat.

Jak z výše popsaného vyplývá, k tomu, aby podobnou studii bylo možné provést, je zapotřebí mít k dispozici materiál, který je do jisté míry kontrolovaný. To sice ve forenzní praxi lze napodobit (srovnávací nahrávka se může pořídit se stejným jazykovým obsahem jako sporná), ale metou forenzního výzkumu je nalézt takový parametr vystihující daného mluvčího, jaký bude robustní vůči různým typům vnějších okolností.

7.4 Statické deskriptory F0

Experimentální část této práce tedy uvádí kromě čteného i spontánní řečový materiál, v němž se zaměříme na statické ukazatele F0. První statistickou metodou, která popisuje tendence mluvčího, je dlouhodobá distribuce hodnot nebo histogram hodnot (pro formanty anglicky LTF – *Long-term formant distribution*, pro F0 LTF_0). Ukazuje, kolikrát se která hodnota F0 v signálu vyskytuje, v grafu, který vykazuje známky normálního rozdělení (největší hustota hodnot je kolem průměru a ve vzdálenosti dvou směrodatných odchylek na každou stranu od průměru). Hodnoty můžou být řazeny do různě velkých intervalů (*bins*), ale pro jemnější vyhlazení je vhodné volit 3 Hz.

Jak píše Skarnitzl (2014: 56), „pro exaktní porovnávání distribucí F0 se běžně používá několik ukazatelů, které se někdy nazývají momenty. […] Kromě aritmetického průměru a směrodatné odchylky distribuce je to sešikmení distribuce (anglicky *skewness*), která vypovídá o symetričnosti distribuce kolem průměru, a špičatost (*kurtosis*), která popisuje, zda jsou hodnoty soustředěny úzce kolem průměru a distribuce je špičatá, nebo zda jsou široce rozprostřené v distribuci plošší."

Dlouhodobá distribuce F0 nám tedy poskytuje určitou představu o dlouhodobých parametrech základní frekvence. Tyto parametry se počítají z té samé celkové sady dat od jednoho mluvčího a můžeme si je rozdělit na střední hodnoty a na ukazatele variability. V experimentální části se budeme zabývat třemi parametry od každého druhu: v rámci středních hodnot půjde o průměr, medián a základní hladinu, v rámci ukazatelů variability o rozpětí F0, rozpětí mezi 10. a 90. percentilem a směrodatnou odchylku.

Kromě porovnání těchto ukazatelů napříč mluvčími a dvěma řečovými styly provedeme také stručnou sondu do stabilizace průměru F0 a základní hladiny. Tato sonda bude volně založená na studii Jana Volína (2007a), který chtěl na čteném materiálu zjistit, kolik sekund řečových dat je potřeba mít, než se údaje o základní frekvenci stabilizují. Odpověď na takové výzkumné otázky může mít přesah do forenzní praxe.

8 Experimentální část

8.1 Metoda

Pro potřeby experimentální části jsme pořídili studiové nahrávky 8 mluvčích ve věkovém rozmezí 20-30 let. Tuto věkovou skupinu jsme zvolili proto, aby náš vzorek byl reprezentativní ve smyslu forenzní praxe. Nejprve jsme účastníky výzkumu požádali, aby přečetli výběr rozhlasových zpráv. Ty později posloužily jako tematický podklad pro spontánní rozhovor, který trval 30 až 45 minut. Všechny nahrávky byly pořízeny ve zvukovém studiu Fonetického ústavu FF UK při vzorkovací frekvenci 32 kHz a 16bitové hloubce; záznam byl proveden přímo do zvukové karty počítače, použit byl kondenzátorový mikrofon AKG C4500 B-BC.

Nejprve jsme zpracovali nahrávky spontánních dialogů. Rozřezali jsme je na repliky (tedy souvislé a experimentátorem nepřerušené úseky promluvy) a tyto repliky podrobili extrakci hodnot F0 v 5ms intervalech pomocí autokorelačního nástroje *pitch tracker* v Praatu (Boersma a Weenink, 2015; Boersma, 1993). Při extrakci byla ponechána standardní nastavení, jen horní hranice pro detekci F0 byla stanovena na 350 Hz, jelikož se jednalo o mužské hlasy. Od každého mluvčího jsme k analýze použili 100.000 těchto hodnot, což odpovídá přibližně osmi minutám znělého signálu.

Cílem prvního experimentu prováděného na spontánním materiálu bylo zjistit, zda a nakolik se mluvčí liší ve zvolených ukazatelích středních hodnot týkajících se základní frekvence – tedy průměru, mediánu a také výše zmíněné základní hladiny (ZH), kterou sice nelze – striktně vzato podle způsobu výpočtu – považovat za střední hodnotu, ale patří mezi deskriptory stability základní frekvence.

Druhý experiment jsme zaměřili na ukazatele variability ve spontánním materiálu, konkrétně rozpětí F0, směrodatnou odchylku a hodnoty mezi 10. až 90. percentilem. Zjištěné výsledky jsme průběžně ověřovali pomocí poslechu, abychom upřesnili, jak se naměřené rozdíly manifestují v tom, jak mluvčí znějí.

Následoval oddíl, který měl dosavadní zjištění konfrontovat se stejným druhem analýz na čteném materiálu. K tomuto účelu jsme použili od každého druhu diskurzu 10 tisíc hodnot na mluvčího, aby byla zajištěna lepší srovnatelnost mezi mluvními styly.

Napříč všemi experimenty jsme prováděli analýzu dlouhodobých distribucí F0, aby bylo možné kontrolovat rozdělení hodnot F0 daného mluvčího ve srovnání jak s ostatními (hodnocení interindividuálních rozdílů), tak se sebou samými (intraindividuální rozdíly).

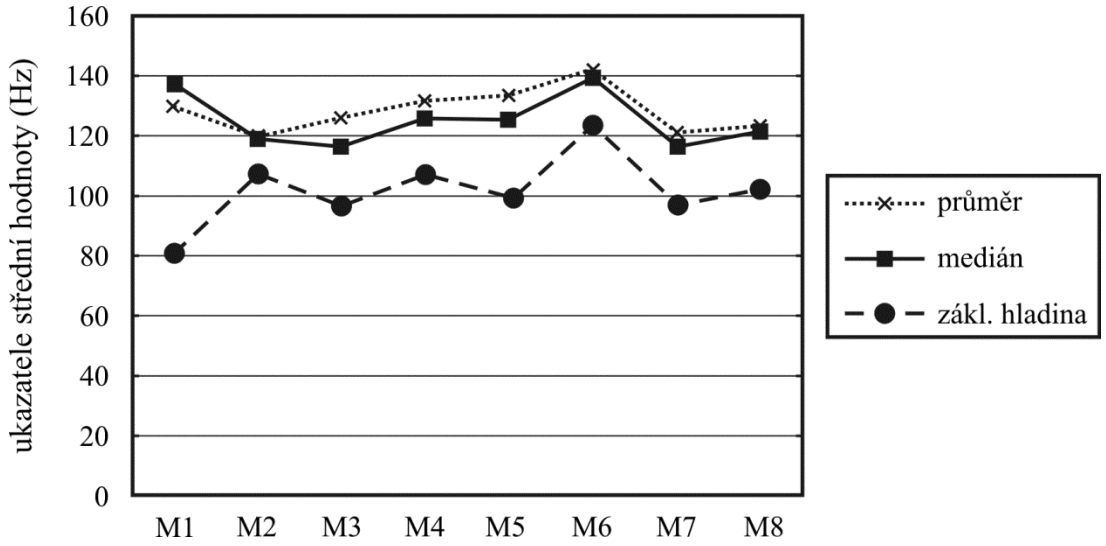
Na závěr jsme provedli krátkou sondu, která měla za úkol ověřit výsledky studie Jana Volína (2007a) týkající se stabilizace deskriptorů F0, přesněji řečeno množství dat, které je k dosažení spolehlivých a přesně vypovídajících ukazatelů potřeba, což je z forenzního hlediska vysoce relevantní. Rozdíl našeho přístupu oproti zmíněné studii spočívá v tom, že jsme použili spontánní materiál místo čteného, což opět dává prostor zajímavému srovnání těchto stylů.

Při této analýze jsme se zaměřili na průměr F0 a základní hladinu (ZH), u níž jsme na základě výsledků studie Lindha a Erikssona (2007) i na základě vlastních dosavadních zjištění předpokládali nižší variabilitu, a tedy rychlejší stabilizaci. Postupovali jsme podobně jako Volín (2007a): krokově jsme přičítali průměrné F0 a ZH každé následující repliky a sledovali, jak se mění celkové průměry těchto ukazatelů.

Výsledky všech avizovaných experimentů představíme a podrobněji rozebereme v následujícím oddíle. Struktura a prezentované výsledky výzkumu vycházejí z kapitoly Skarnitzl a Hývlová (2014), s tím, že v této diplomové práci je řečový materiál obohacen o čtené nahrávky (viz výše), což nám umožnilo zaměřit se na vliv druhu diskurzu na chování základní frekvence.

8.2 Analýza spontánního materiálu

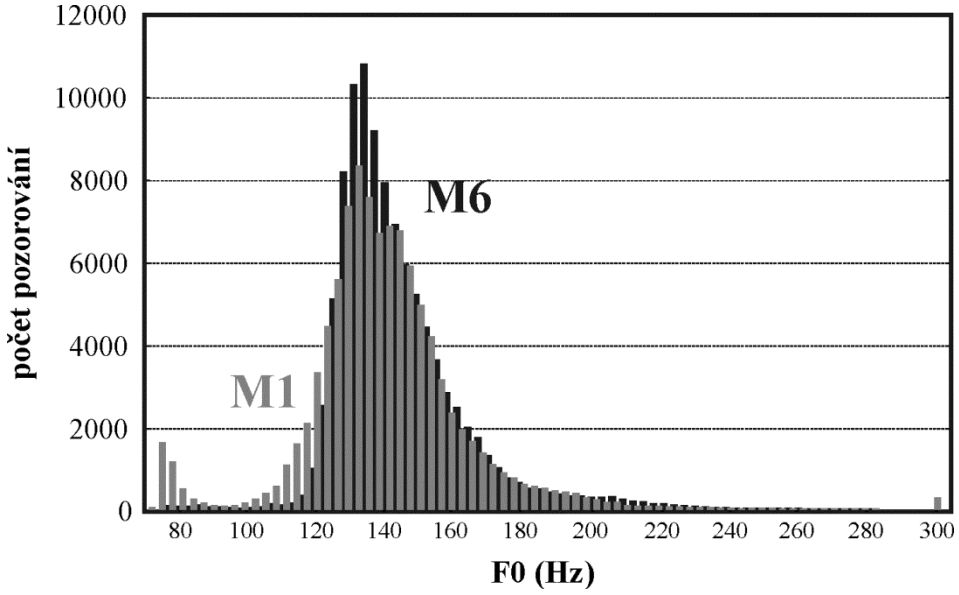
8.2.1 Výsledky I: Střední hodnoty u spontánního materiálu



Obr. 8.1. Střední hodnoty F0 analyzované u mluvčích M1-M8.

Podíváme-li se tedy nejprve na základní hladiny a ukazatele středních hodnot, z obrázku 8.1 je patrné, že jsou mezi nimi u mluvčích podobné odstupy. To naznačuje, že mluvčí mají podobné počty nejnižších hodnot. Výjimku tvoří mluvčí M1, jehož ZH leží výrazně níže než průměr a medián. Snížená je u tohoto mluvčího zřejmě proto, že se u něj často vyskytuje třepená (ale autokorelací stále detekovatelná) fonace, což se projevuje ve zvýšeném množství hodnot v rozmezí 70-90 Hz. Tuto idiosynkratickou vlastnost mluvčího M1 lze dobře vidět v jeho histogramu v obrázku 8.2 či 8.5 níže. Kumulace hodnot F0 nižších než 100 Hz způsobuje, že osmý percentil, který základní hladině odpovídá (viz oddíl 6.6), leží v nižších frekvencích než u ostatních mluvčích.

Srovnáme-li jej s mluvčím M6, který rovněž svými středními údaji vybočuje ze skupiny, ten má sice stejně jako M1 charakteristický vysoký hlas, ale bez třepené fonace. Dlouhodobé distribuce F0 na obrázku 8.2 ilustrují, do jaké míry se hodnoty F0 u mluvčích M1 a M6 překrývají. Třepenou fonaci, která je jedním z mála rozdílů mezi mluvčími a není okamžitě zachytitelná pouhým poslechem, tedy vystihuje ukazatel základní hladiny.



Obr. 8.2. Distribuce F0 pro mluvčí M1 a M6.

Naše pozorování se shoduje s poznatkem kolektivu autorů Hudson *et al.* (2007: 1810), kteří rovněž přišli na to, že modus tří jimi analyzovaných mluvčích se pohyboval kolem 50-60 Hz. Pohled na histogramy jim také odhalil, že třepený hlas vytváří u mluvčího vedlejší vrcholek v distribuci F0. Na základě toho modus těchto mluvčích upravili. Jejich další závěry týkající se modu se týkaly například toho, že tento ukazatel svou přesností závisí na velikosti intervalů v histogramu (*bins*), do nichž se hodnoty F0 třídí. Pro svou studii zvolili velikost 5 Hz, naše distribuce využívají ještě vyšší jemnost – 3 Hz.

Na tomto místě je vhodné poznamenat, že autoři považují modus za lepší ukazatel než průměr, jelikož průměr vytahují nahoru nahodilé odlehlé hodnoty ve vyšších frekvencích – přitom je přirozené, že F0 se u mluvčího pohybuje blízko spodní hranici hlasového rozsahu. Když u mluvčích odečetli modus od průměru, ve většině případů vycházelo kladné číslo mezi 1 – 8 Hz (tamtéž). Menší citlivost na odlehlé hodnoty uvádějí jako podpůrný argument i Lindh a Erikson (2007; srovnej i s Lindh, 2006), ale pro medián. Z pohledu na obrázek 8.1 je jasné, že medián vhodněji kopíruje chování základní hladiny napříč mluvčími než aritmetický průměr.

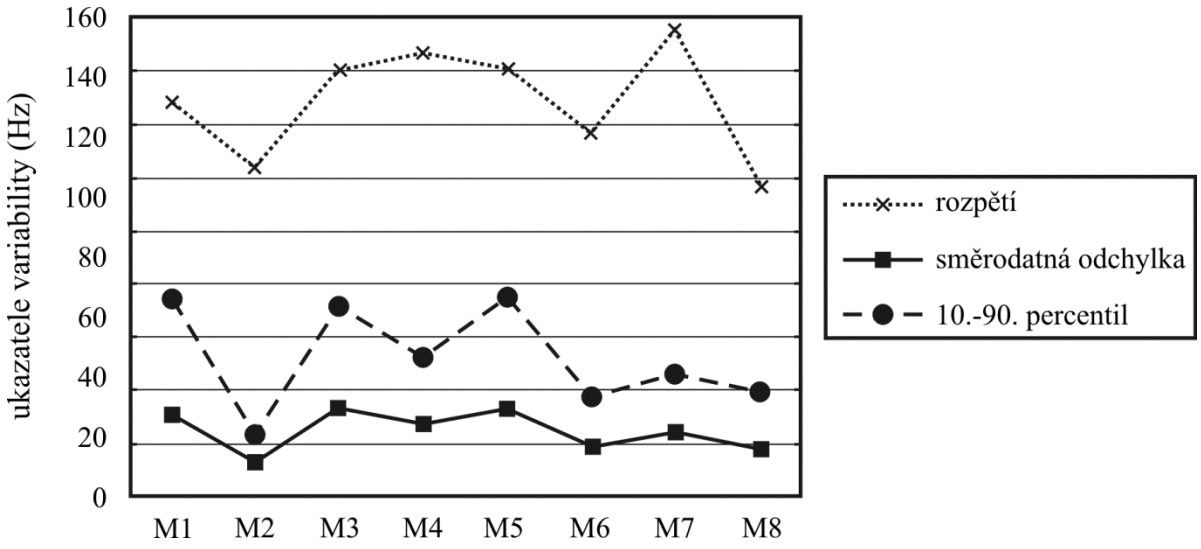
Vrátíme-li se k naší diskusi třepené fonace, lze konstatovat, že můžeme s ní spojené výsledky interpretovat jako doporučení pro forenzní praxi, co se týká středních hodnot: jako výhodné se jeví užívání průměru v kombinaci s modem či mediánem, nikoli zvlášť. Dále pak je užitečné přihlédnout k základní hladině a možným diskrepancím mezi ní a ostatními

ukazateli, jelikož tyto nesrovnalosti mohou prozradit forenzně využitelnou odlišnost mezi mluvčími.

8.2.2 Výsledky II: Ukazatele variability u spontánního materiálu

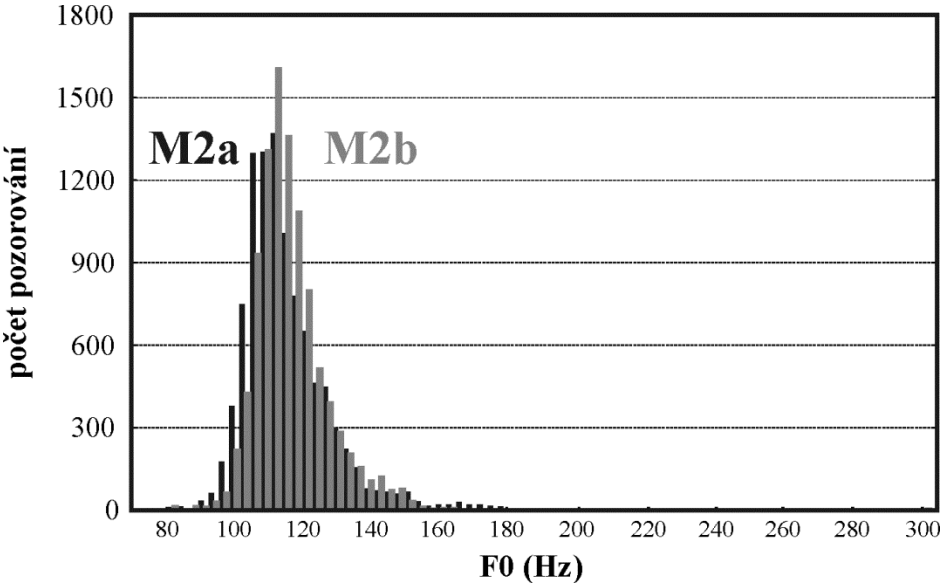
Ukazatele variability můžeme porovnat na obr. 8.3. Vidíme, že samotné rozpětí hodnot F0 (minimální hodnota odečtená od maximální) se nechová vždy stejně jako percentilové rozpětí, u kterého bylo před tímto výpočtem odebráno 10% nejvyšších a 10% nejnižších hodnot. U mluvčích M4 a M7 je odlišné chování rozpětí obzvláště markantní, ale lze ho vysvětlit zřejmě vyšší mírou chybné extrakce F0 (oktávových skoků, viz oddíl 7.1), která je pro materiál přirozená. Připomeňme, že manuální úprava extrahovaných hodnot ve forenzním kontextu není realistická ani žádoucí, protože ubírá celé proceduře na efektivitě.

Užití 10.–90. percentilu poskytuje o frekvenčním rozpětí tedy lepší představu. Je zřejmé, že percentilové rozpětí mluvčích M1, M3 a M5 je vysoké. U mluvčího M1 je to alespoň částečně dáno zmíněnou třepenou fonací, u mluvčích M3 a M5 lze předpokládat z intonačního hlediska skutečně variabilnější projev. V jejich případě však může hrát roli i přítomnost „cizích“ frekvencí, zejména překryv se vstupy moderátorky rozhovoru. I zde platí, že s výskytem takových vnějších zvuků je třeba ve forenzní praxi počítat.



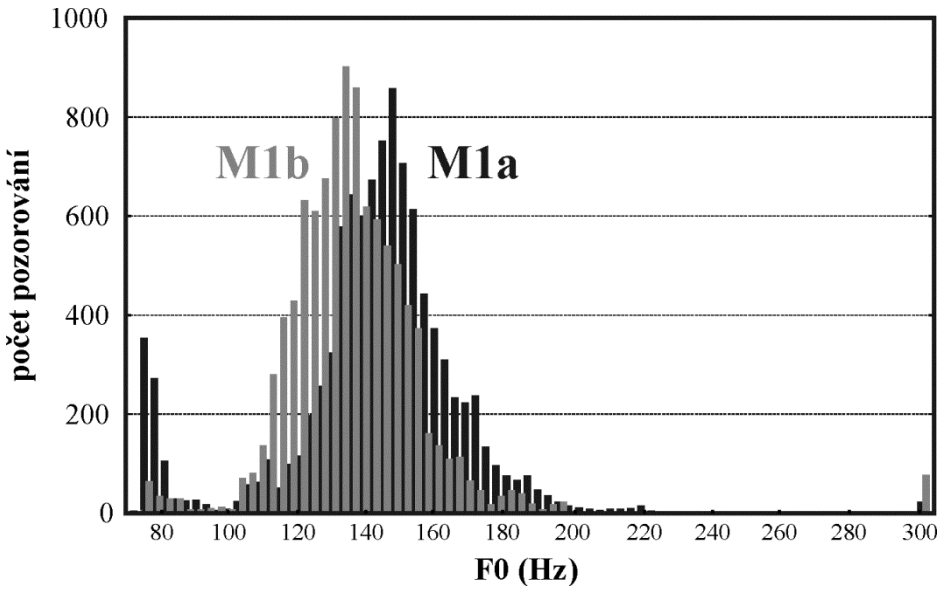
Obr. 8.3. Ukazatele variability F0 analyzované u mluvčích M1-M8.

Za bližší zmínku stojí mluvčí M2, který má na první pohled malý odstup průměru od ZH (obr. 8.1) a zároveň nízkou variabilitu (obr. 8.3), podle čehož se dá očekávat poměrně kompaktní soubor dat. Obrázek 8.4 – na němž jsou porovnány distribuce F0 mluvčího M2 z první a druhé poloviny celého dialogu – toto očekávání potvrzuje. Zároveň je z obrázku zřejmé, že tento mluvčí svůj projev (alespoň z hlediska F0) v průběhu času výrazně neměnil. Na závěr dodejme, že při poslechu mluvčí M2 skutečně zní ve srovnání s ostatními mluvčími značně monotónně.



Obr. 8.4. Srovnání distribuce F0 mluvčího M2 v první (M2a) a druhé (M2b) polovině nahrávacího času.

Při detailnějším pohledu na variabilnějšího mluvčího M1 vychází najevo, že nevyniká mezi ostatními pouze díky třepené fonaci, ale i širokému frekvenčnímu rozdělení hodnot. Jelikož první a druhá polovina dat koresponduje s první a druhou polovinou času rozhovoru, obrázek 8.5 naznačuje, že se F0 tohoto mluvčího s postupem času snižovala. Podobný efekt by ale ve forenzních nahrávkách zřejmě nebyl pozorovatelný, protože nebývá k dispozici tolik materiálu; tato analýza je založena na dlouhých nahrávkách.

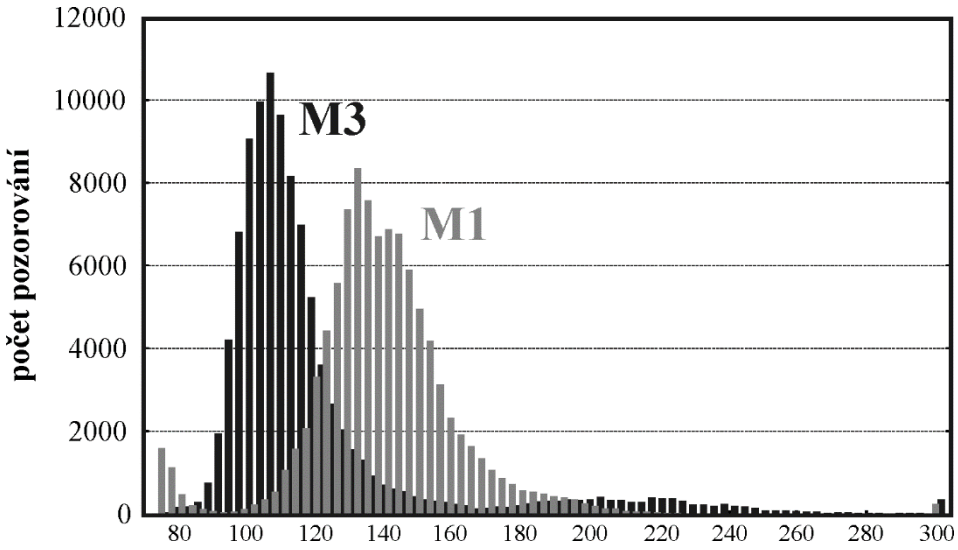


Obr. 8.5. Srovnání distribuce F0 mluvčího M1 v první (M1a) a druhé (M1b) polovině nahrávacího času.

8.2.3 Intraindividuální variabilita: spontánní materiál

Na tomto místě je vhodné zastavit se nad intraindividuální variabilitou. Kolektiv autorů Skarnitzl et al., (2014) na dlouhodobých distribucích vokalických formantů pomocí Kolmogorov-Smirnovova testu ukázal, že tyto distribuce vykazují dostatečně nízkou míru variability v rámci mluvčího. Chtěli jsme tedy zjistit, zda za podobně stabilní můžeme považovat i dlouhodobé distribuce základní frekvence. Porovnání distribucí F0 ve dvou časových polovinách (jak ukazují obrázky 8.4 a 8.5) pomocí Kolmogorov-Smirnovova testu přineslo ve všech osmi srovnáních vysoce významný rozdíl ($p < 0{,}001$), což bohužel značí variabilitu v rámci daného mluvčího. Zatímco v případě formantů nejpříznivější výsledky přineslo systematické rozdělení, kdy byly porovnávány liché vzorky nahrávky se sudými (rozdíl získaný K-S testem byl u všech srovnání nevýznamný), u LTF0 jsme ani touto metodou takového výsledku nedosáhli. Jen dvě z osmi intraindividuálních srovnání nedosáhla statistické významnosti ($p > 0{,}1$), u dvou srovnání je $p < 0{,}05$ a u zbývajících čtyř dokonce $p < 0{,}005$. Autoři oddíl 2.3.1. uzavírají s tím, že K-S test je velmi citlivý. U distribucí F0 v rámci mluvčích se to projevilo dvojnásob. Můžeme tím pádem říci, že tento druh analýzy není vhodný pro odlišení těch mluvčích, kteří inkonzistence skutečně vykazují, od těch, jejichž data jsou průběžně kompaktní.

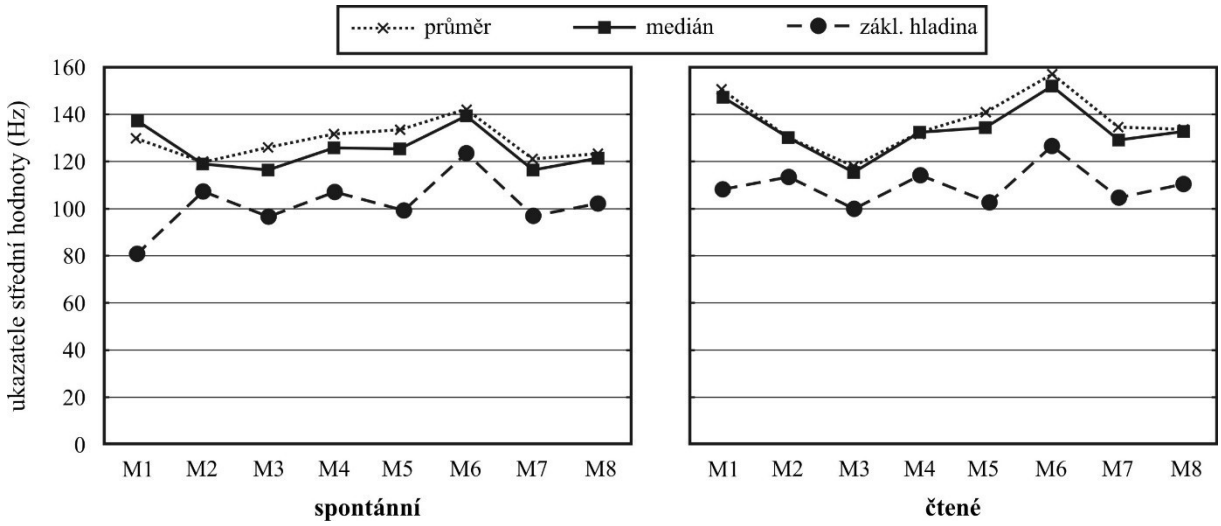
Na závěr komentáře ke spontánnímu materiálu se vraťme k užitečnosti kombinování více ukazatelů F0 pro rozlišení mluvčích. Kdybychom například vycházeli pouze z průměru a směrodatné odchylky (obr. 8.1 a 8.3), mohli bychom mluvčí M1 a M3 považovat za srovnatelné a očekávat, že se jejich data budou překrývat. Jak ale naznačuje obrázek 8.6, LTF0 obou mluvčích se značně liší. Již jsme hovořili o tom, že se tito mluvčí odlišují v základní hladině kvůli vyšší přítomnosti třepené fonace u mluvčího M1. Je tedy prospěšné si při forenzně-fonetické analýze zobrazovat ukazatelů více a mít na paměti jejich odlišnou spolehlivost a robustnost. Zároveň je vhodné nezapomínat také na dlouhodobé distribuce hodnot, které mohou leckdy pomoci ozřejmit rozdílnosti či shody, jež se projeví při srovnávání ukazatelů napříč mluvčími.



Obr. 8.6. Distribuce hodnot F0 pro mluvčí M1 a M3.

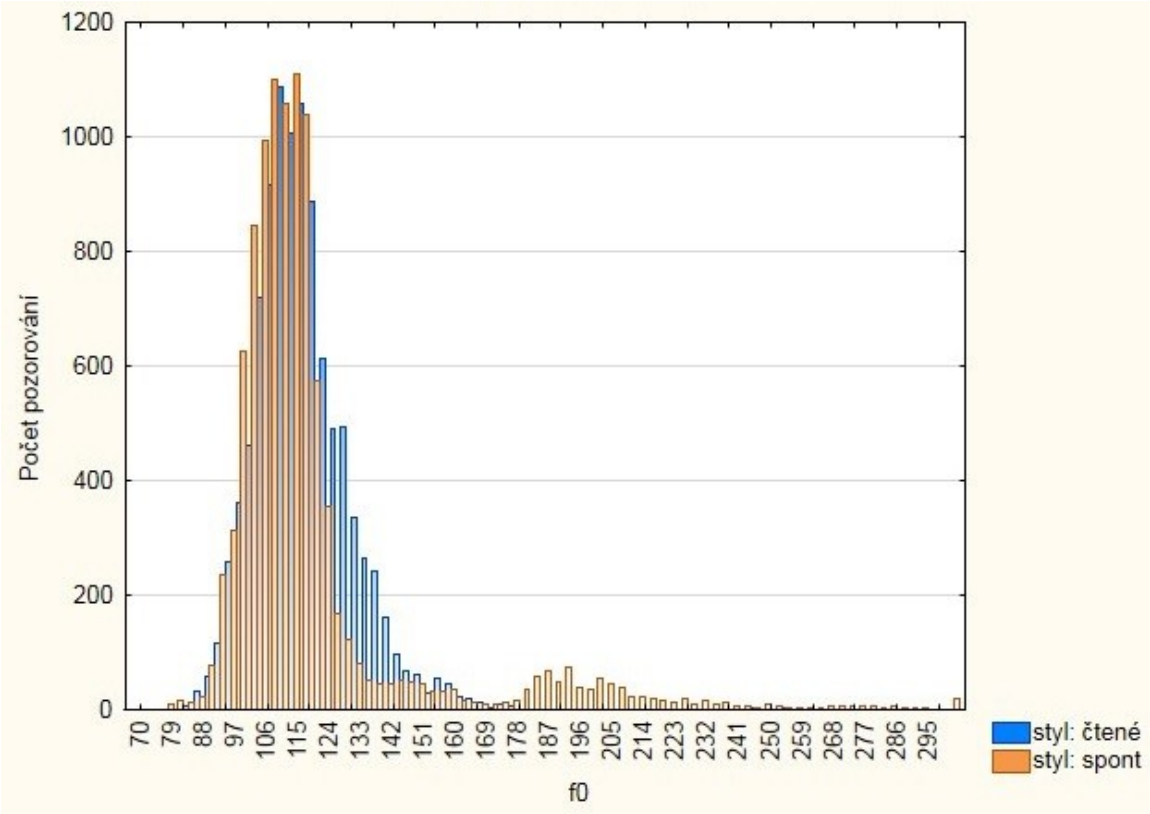
8.3 Závislost na mluvním stylu: Porovnání se čteným materiálem

Abychom ověřili výsledky studií zmiňovaných v teoretickém úvodu (Hollien, 1997; Skarnitzl a Vaňková, 2015) a zároveň učinili srovnání s výstupy analýz na spontánních dialozích, podrobili jsme čtený materiál našich 8 mluvčích rovněž analýze středních hodnot a ukazatelů variability. Hypotézami bylo, že střední hlasová poloha bude u čteného materiálu vyšší než u spontánního a také že rozptyl hodnot bude při čtení vyšší než v rámci spontánního dialogu (obojí vychází ze srovnání s výsledky Skarnitzl a Vaňková, 2015; viz oddíl 8.4).



Obr. 8.7. Střední hodnoty analyzované u spontánního materiálu (vlevo) a čteného materiálu (vpravo).

Výsledky v obrázku 8.7 nasvědčují tomu, že průměrná střední hlasová poloha vyšší skutečně je, a to téměř až o 20 Hz (nejvýrazněji se to projevuje u mluvčího M6). Průměr i medián jsou v případě čteného materiálu konzistentně vyšší a lépe se spolu shodují. Výjimku tvoří mluvčí M3, u nějž nejen že se neposunul medián, ale dokonce se u něj snížil průměr. Toto lze vysvětlit z bližšího pohledu na distribuce hodnot M3, spontánních i čtených dohromady. Obě distribuce se značně překrývají, ale u jeho spontánních hodnot se vyskytuje velké množství *outliers*. Znovu se zde potvrzuje, že medián (který je odlehlými hodnotami modifikován méně) lépe postihuje realitu.



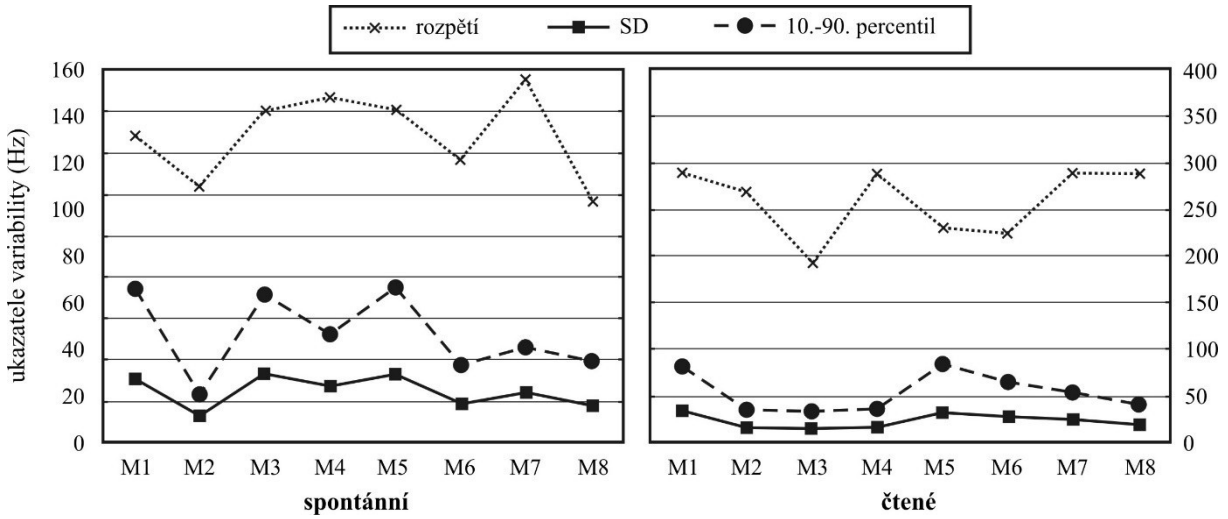
Obr. 8.8. Distribuce mluvčího M3, dva mluvní styly: čtený a spontánní.

Z obrázku 8.7 vyplývá ještě jedno důležité zjištění, a sice že základní hladina nepodléhá ve čteném materiálu tak výrazným zvýšením jako průměr a medián, srovnáváme-li tyto hodnoty se spontánním materiálem. Zejména u mluvčích M3, M5 a M6 jsou posuny jen zanedbatelné. Opět jedinou výjimkou je mluvčí M1, u nějž se základní hladina posunuly výše na úroveň běžnou pro ostatní mluvčí. Lze to přičítat tomu, že se u jeho čteného projevu daleko méně setkáváme s výše zmiňovanou třepenou fonací na konci fráze (podrobnější intraindividuální srovnání napříč mluvními styly provedeme v 8.6 níže). Chování základní hladiny v našem srovnání je převážně v souladu se zjištěním autorů Lindh a Eriksson (2007) ohledně robustnosti tohoto ukazatele vůči vlivům živosti projevu.

Rozdíly v živosti obou druhů projevu můžeme dobře posoudit na ukazatelích variability, které srovnává obrázek 8.9. Zatímco u směrodatné odchylky opět vidíme od mluvčího k mluvčímu pouze mírné změny, které odrážejí špičaté či plošší rozložení hodnot kolem

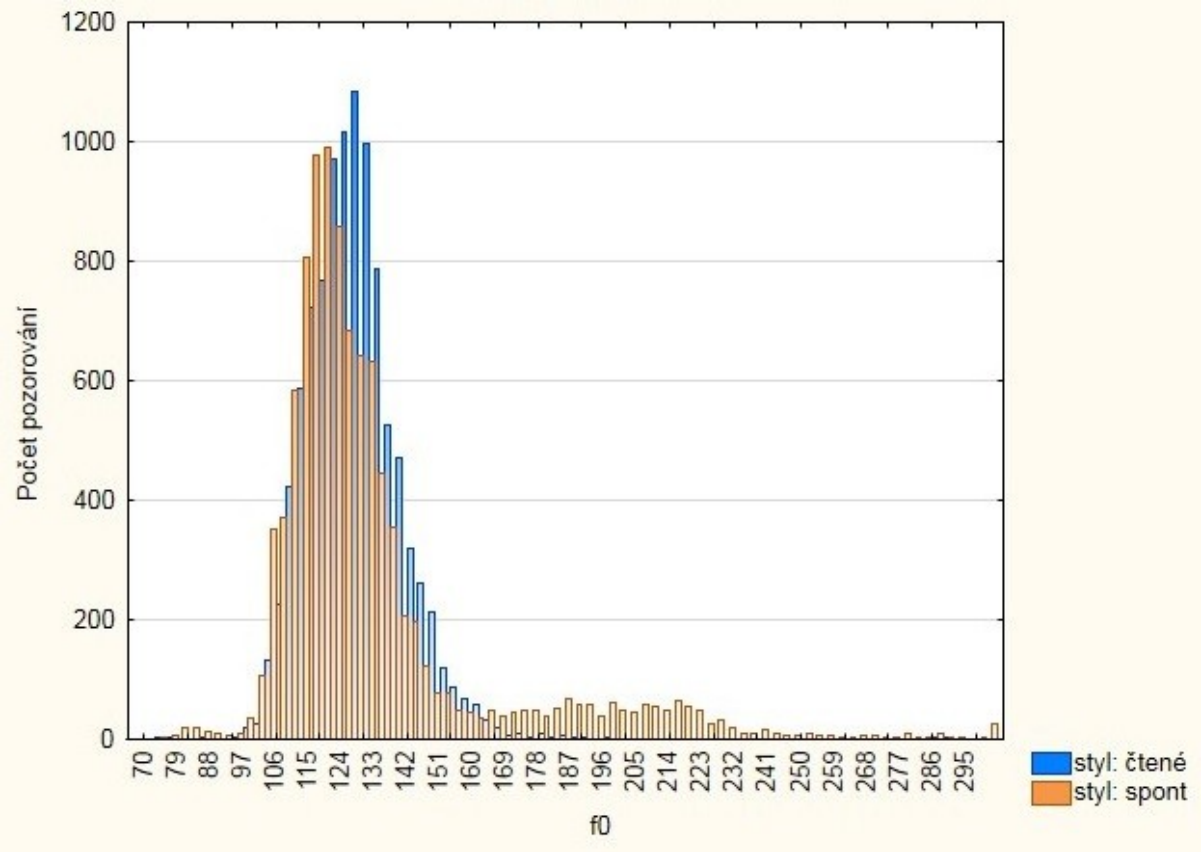
průměru (u ploššího je řeč zejména o mluvčích M1, M5 a M6 – viz obrázek 8.13 dále), zajímavější srovnání poskytují druhé dva ukazatele.

Museli jsme upravit svislé osy s jednotkou Hz, protože rozpětí hodnot F0 bylo u čteného materiálu markantně větší. To mluví ve prospěch hypotézy, že rozptyl hodnot a tedy hlasová modulace je u čtení textu výraznější. V našem případě můžeme vyloučit, že by to bylo způsobeno přítomností pasáží přímé řeči v textu – jednalo se o přepisy rozhlasových zpráv.



Obr. 8.9. Ukazatele variability analyzované u spontánního materiálu (vlevo) a čteného materiálu (vpravo).

Při bližším pohledu na chování běžného rozpětí (tedy rozdílu mezi maximální a minimální hodnotou v souboru dat) v obrázku 8.9 se však znovu potvrzuje, že tento ukazatel nelze brát příliš vážně, jelikož s odpovídající hodnotou percentilového rozpětí u daného mluvčího koresponduje ještě méně, než jak tomu je u spontánního materiálu. Přicházíme tak znovu ke zjištění z oddílu 8.2.2 (Výsledky II). Běžné rozpětí dává představu, že variabilita u čteného materiálu je skutečně výrazně vyšší (ve většině případů nad 100 Hz rozdíl), ale percentilové rozpětí tuto představu upravuje. Např. u mluvčího M4 je u čteného materiálu rozpětí o 120 Hz širší, kdežto po oříznutí 10. a 90. percentilem vychází najevo, že je ve skutečnosti jeho variabilita o několik Hz nižší. Původ této diskrepance opět názorně vysvětlují odlehlé hodnoty v histogramu mluvčího M4 (obr. 8.10).

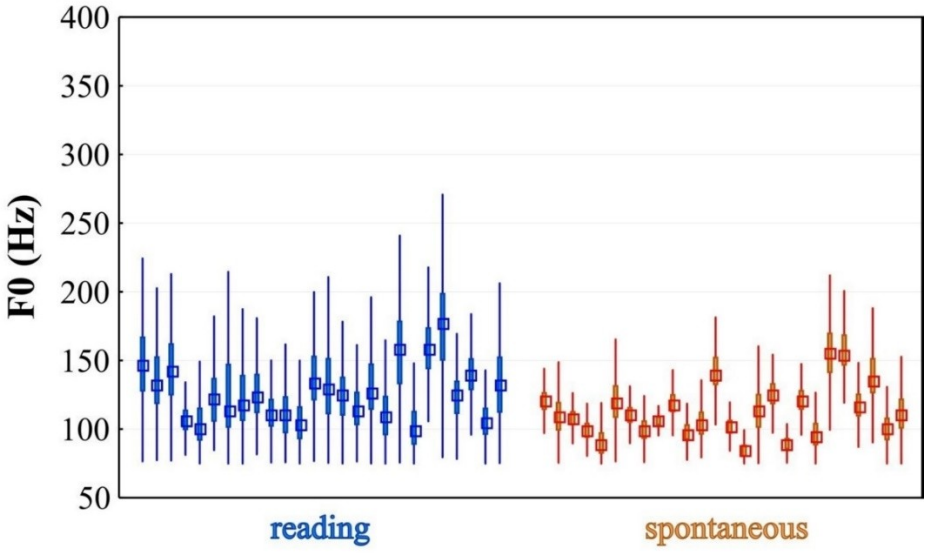


Obr. 8.10. Distribuce mluvčího M4, dva mluvní styly: čtený a spontánní.

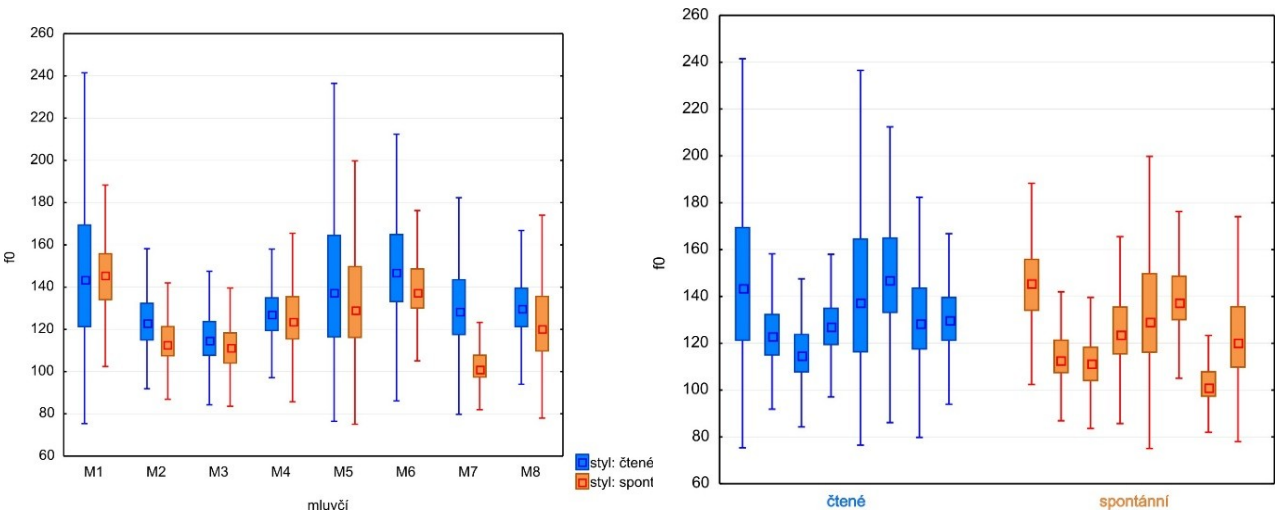
Co se týče skutečného rozdílu v chování mluvčích od spontánního ke čtenému stylu, největší nárůst variability zaznamenal mluvčí M6 (27 Hz v percentilovém rozpětí), naopak u mluvčího M3 vidíme propad variability o 40 Hz, a to i přesto, že percentilové rozpětí umožňuje korekci odlehlých hodnot, které u spontánního materiálu tohoto mluvčího pozorujeme. To je oproti mírným nárůstům variability u ostatních mluvčích výrazný rozdíl. Jsou to právě tyto podstatnější odlišnosti v řečovém chování, kvůli nimž má smysl srovnání napříč mluvními styly dělat a uplatňovat i ve forenzní praxi.

8.4 Srovnání s výsledky Skarnitzl, Vaňková (2015): krabicové grafy

Jak jsme již avizovali, jedním ze záměrů této experimentální části je i provést srovnání s výsledky studie autorů Skarnitzl a Vaňková (2015). Ti velmi podobně zkoumali 26 mluvčích ve věkovém rozpětí 20 – 45 let a ukazatele vztahující se k jejich základní frekvenci v rámci 3 mluvních stylů: čteného stylu, spontánního stylu a maskovaného hlasu. V rámci svých výsledků sestavili krabicové grafy, které nabízejí srozumitelné vizuální srovnání mediánů, rozsahů hodnot mezi 25.-75. percentilem a rozpětí, která neobsahují odlehlé hodnoty (viz obrázek 8.11). Pro účely srovnání jsme použili data ze čteného a spontánního materiálu.



Obr. 8.11. Krabicové grafy popisující čtený a spontánní materiál ze studie Skarnitzl a Vaňková (2015)**.**

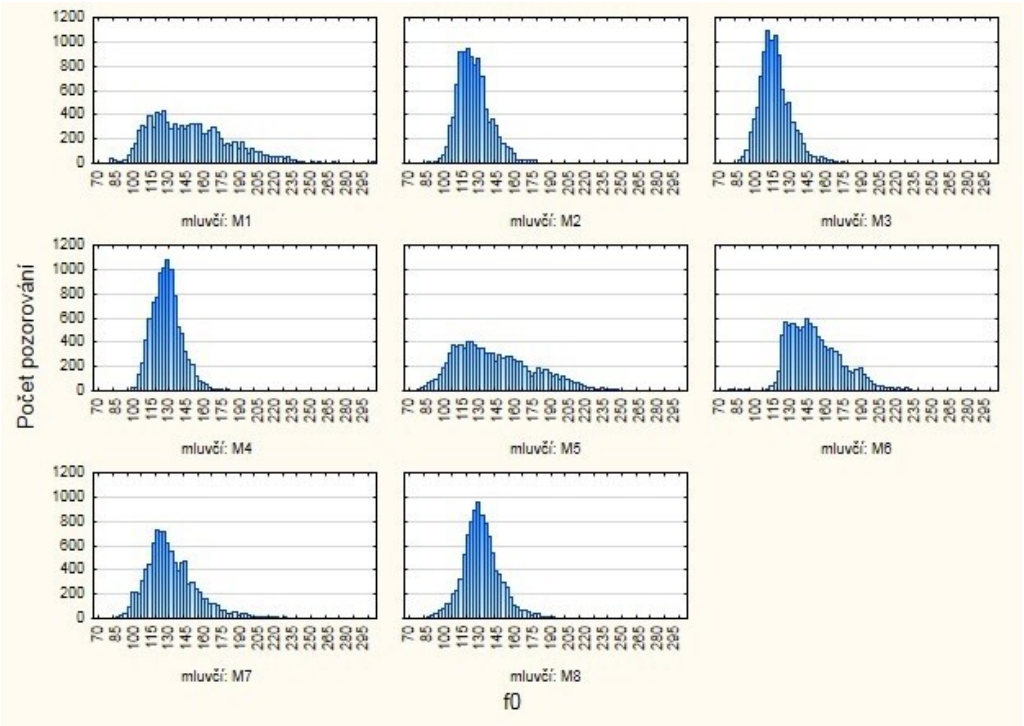


Obr. 8.12. Krabicové grafy našeho čteného a spontánního materiálu, uvedené pro srovnání.

Jak je z obrázků patrné, oboje výsledky vykazují stejné tendence, a to i přesto, že se u Skarnitzla a Vaňkové (2015) jednalo o jiný typ čteného materiálu (pohádkový druh textu obsahující přímou řeč). Mediány se v obou případech pohybují kolem 130-140 Hz. U těchto krabicových grafů jsou lépe pozorovatelná zejména větší rozpětí hodnot F0 u čteného textu. Kromě toho z nich lze snadno vyčíst rozdělení dat v rámci jednotlivých mluvčích, jelikož výrazná oblast ohraničená 25. a 75. percentilem reflektuje soustředěnost hodnot kolem průměru (krátké střední bloky) nebo větší rozptýlenost (dlouhé bloky; pro srovnání viz opět obrázek 8.13).

8.5 Interindividuální variabilita: čtený materiál

Podíváme-li se na obrázek 8.13, který znázorňuje distribuce hodnot F0 všech mluvčích u jejich čteného projevu, vhodně doplňuje dosavadní výsledky. Je na něm vidět, že střední hodnoty se skutečně nejčastěji pohybují kolem 130 Hz, zato ve variabilitě se mluvčí odlišují, zvláště pak mluvčí M1, M5 a M6. Jejich odlišná práce s hlasem (která je zřetelná i při poslechovém hodnocení čtených nahrávek) se viditelně projeví na plochosti či špičatosti histogramů.

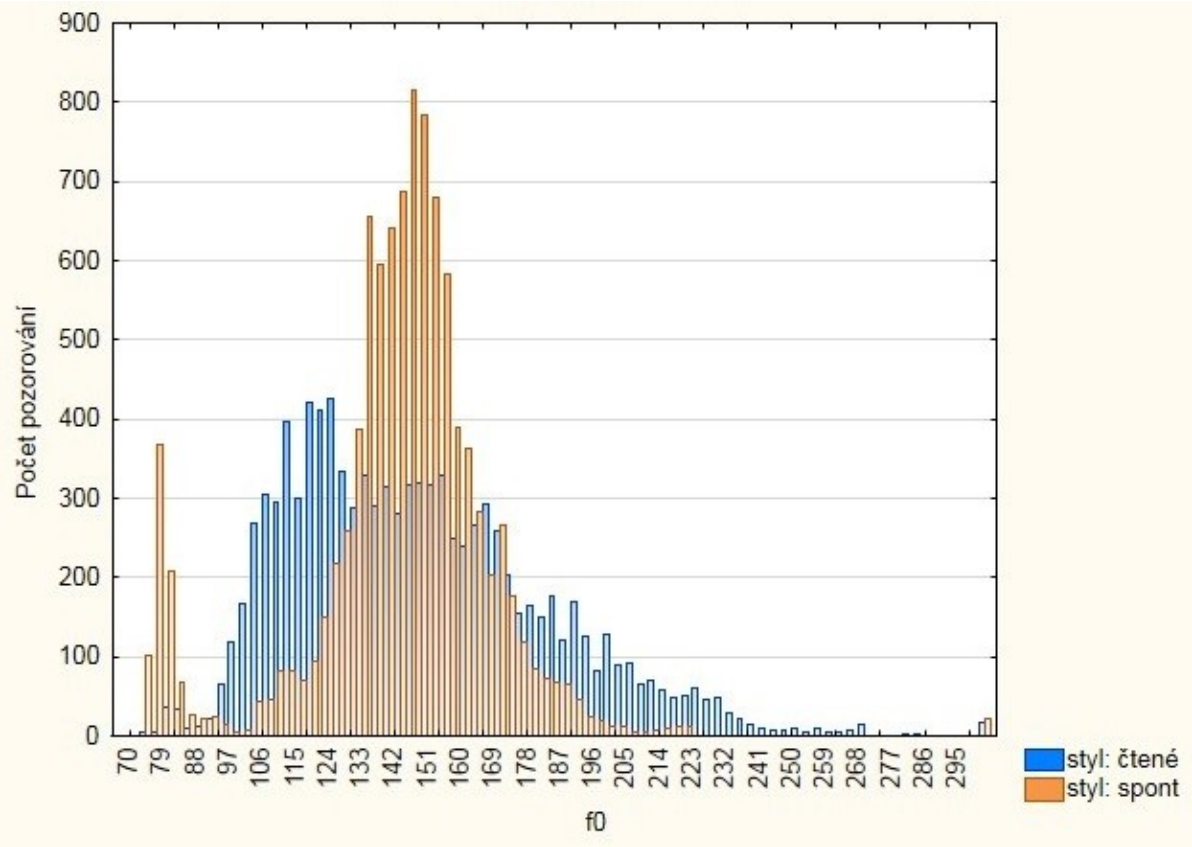


Obr. 8.13. Distribuce hodnot F0 u čteného materiálu pro všech 8 mluvčích.

Variabilitu F0 tedy histogramy hodnot postihují dobře, co však z nich nevyplývá, je hlasová kvalita. Ačkoli by se při pohledu na distribuce mluvčích M2 a M8 mohlo zdát, že jejich hlasy budou znít podobně, poslech to nepotvrzuje. Informace o odlišné hlasové kvalitě tedy v našich statistikách chybí, pokud se právě nejedná o třepenost fonace, jak jsme uvedli v oddíle 8.2.2 (Výsledky I). Takovou informaci by částečně doplnil spektrální sklon, ale ani ten nedokáže zachytit idiolektickou odlišnost mluvčích, rozdílná tempa či různé artikulační návyky, např. větší či menší nosovost. Přesně z toho důvodu by ve forenzní analýze nemělo být opomíjeno ani systematické poslechové hodnocení.

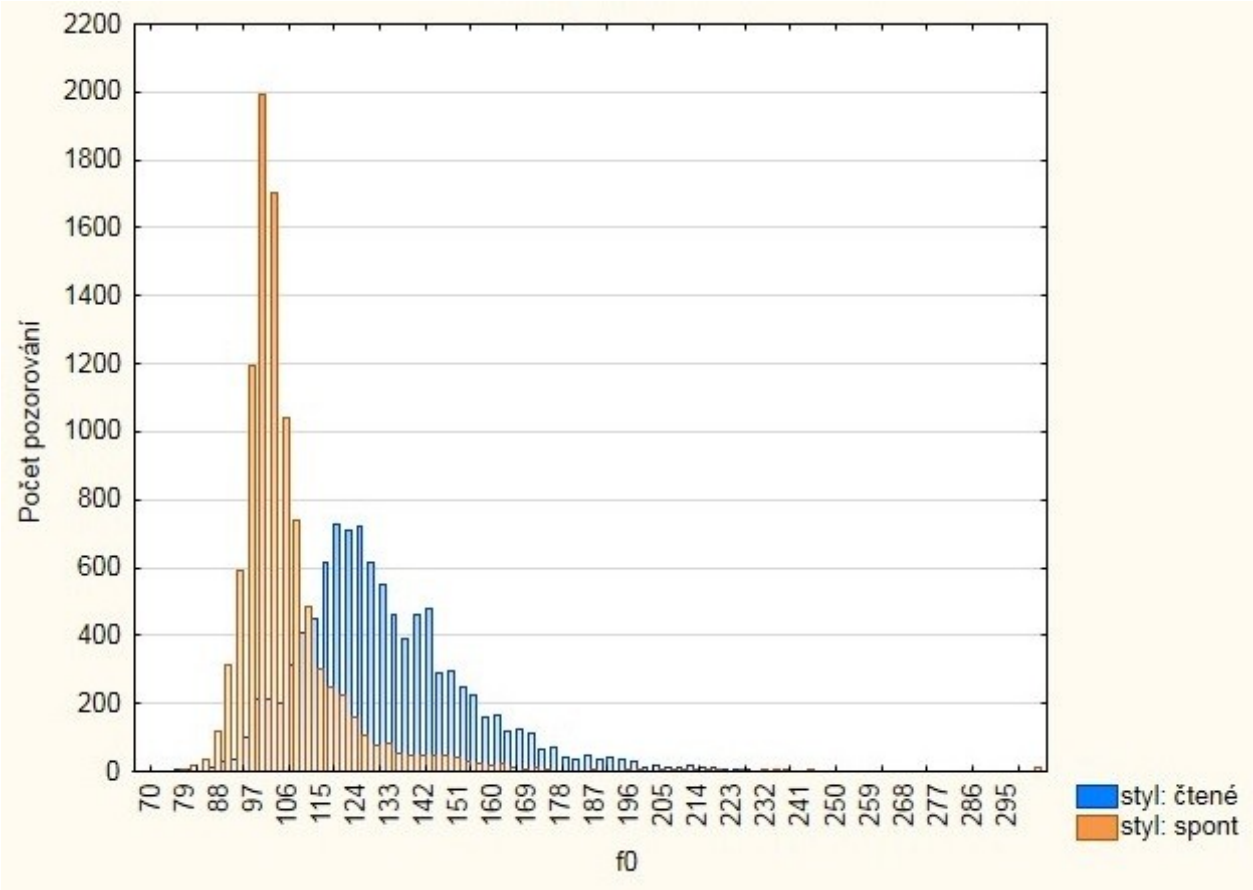
8.6 Intraindividuální variabilita: čtený a spontánní mluvní styl

Podobně jako jsme zkoumali variabilitu v rámci mluvčího pomocí rozdělení jejich datových sad na dvě poloviny, podívejme se nyní na porovnání čteného a mluvního stylu v rámci mluvčího. Variabilita mezi subjekty je ve forenzní fonetice vždy posuzována ve vztahu k variabilitě v rámci jednoho subjektu. Ideální je maximální konzistence ve vlastním projevu, na jejímž pozadí má poté větší váhu odstup hodnot od datové sady jiného mluvčího. V tomto oddíle uvádíme histogramy dvou mluvčích, u nichž dochází k výrazné intraindividuální variabilitě napříč použitými mluvními styly.



Obr. 8.14. Distribuce mluvčího M1, dva mluvní styly: čtený a spontánní.

Mluvčí M1 byl doposud vždy zajímavým subjektem, protože u něj pozorujeme jednak idiosynkratickou třepenou fonaci v nižších frekvencích u spontánních hodnot F0, jednak vysokou variabilitu u čtených hodnot F0, která ho odlišuje od ostatních mluvčích. Obojí z něj činí kandidáta na vysoce spolehlivou rozpoznatelnost v rámci daného mluvního stylu; musíme však vzít v úvahu, že napříč mluvními styly se může jevit jako dva různí mluvčí, přestože oboje hodnoty pocházejí od něj. Je potom otázka, zda ho tato skutečnost diskvalifikuje, či zda rozpoznatelnost od ostatních mluvčích převažuje; my se přikláníme k druhé možnosti, s tím, že se jeví jako prospěšné držet se při srovnávání nahrávek mezi mluvčími vybraného mluvního stylu, ať už čteného, či spontánního.



Obr. 8.15. Distribuce mluvčího M7, dva mluvní styly: čtený a spontánní.

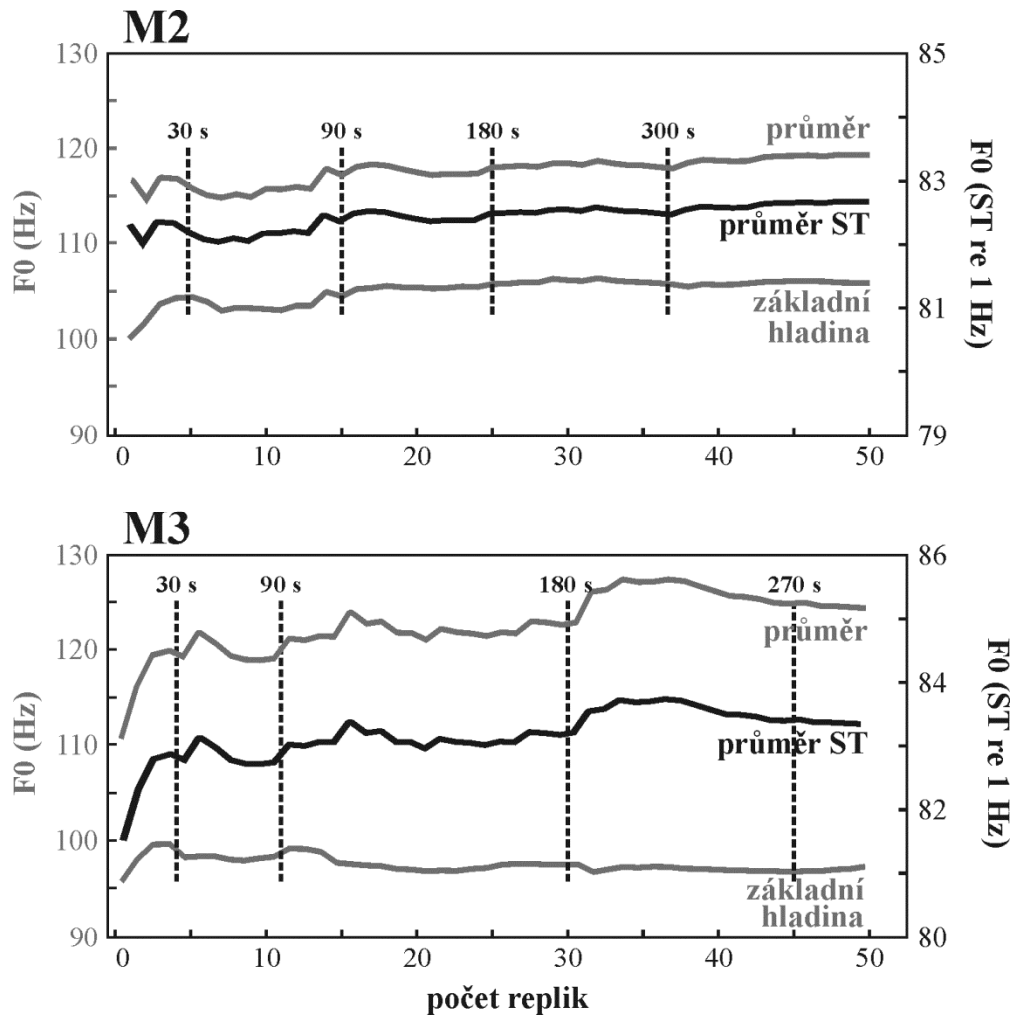
Druhým mluvčím, kterého jsme vybrali pro toto srovnání, je M7. Z obr. 8.15 je patrné, že u svého čteného projevu udržoval vyšší střední hlasovou frekvenci než u spontánního, tedy jeho střední hodnoty se u obou stylů výrazně liší. V jeho případě bychom však rádi odkázali zpět k obrázku 8.7 (střední hodnoty), podle nějž se základní hladina u tohoto mluvčího výrazně ve srovnání obou mluvních stylů nemění. Rozdíl v jejích hodnotách je 7 Hz, podobně jako u všech ostatních. Tato skutečnost hovoří výrazně ve prospěch hypotézy, že základní hladina je ukazatel robustní vůči variabilitě pocházející z odlišných podmínek nahrávání. Zároveň platí, že u mluvčího M1, kterého jsme komentovali výše, tento rozdíl činí 26 Hz. To je však vysvětlitelné již zmiňovanou třepenou fonací přítomnou v nižších frekvencích pouze u jeho spontánních nahrávek. Domníváme se tedy, že je užitečné konsolidovat poznatek o odlišných distribucích napříč mluvními styly s ukazatelem základní hladiny a vytvořit určitou hierarchii těchto deskriptorů z hlediska jejich váhy; výrazný posun základní hladiny by měl ihned

experimentátora nasměrovat k hledání zdroje výrazné intraindividuální variability. Naproti tomu nedokonalý překryv distribucí od jednoho mluvčího nemusí nutně znamenat intraindividuální variabilitu, která by opodstatňovala pochybnost, zda se vůbec jedná o téhož mluvčího.

8.7 Stabilizace průměru F0 a základní hladiny

Dalším cílem bylo také ověřit výsledky studií autorů Edlund a Heldner (2007) a Volín (2007a), kteří se zabývali množstvím řečového materiálu potřebného k přijatelné stabilizaci průměru F0. Autoři studií se shodují na tom, že po 20 sekundách řeči už je statistický průměr F0 jako ukazatel stabilizovaný a spolehlivý – tedy v tom smyslu, že přidávání dalších dat ho již téměř neposouvá. Obě studie však vycházely z čteného textu; my jsme chtěli ověřit, jestli se doba potřebná ke stabilizaci F0 v našem spontánním materiálu bude shodovat.

Obrázek 8.16 ukazuje kumulativní průměr tří parametrů u mluvčích M2 a M3 během prvních 50 replik (s různým trváním). Kromě základní hladiny jsme jako zástupce středních hodnot vybrali průměrnou základní frekvenci vyjádřenou v hertzech a v půltónech; při podobných úlohách se půltónová (logaritmická) stupnice opakovaně ukazuje jako vhodnější (viz např. Nolan, 2003 nebo Volín, 2007). Na obrázku pozorujeme, že u mluvčího M2 se všechny tři ukazatele – průměrná F0 vyjádřená v hertzech a v půltónech i základní hladina – stabilizovaly zhruba po 15 replikách, což v tomto případě znamenalo 90 sekund. U mluvčího M3 došlo ke stabilizaci ZH také po 15 replikách – neboli 102 sekundách – avšak k ustálení jeho průměrné F0 (v hertzech i půltónech) bylo zapotřebí více než 240 sekund. To je v souladu s úvahou, že základní hladina poskytuje spolehlivou představu o průměrném melodickém chování mluvčího na kratším časovém úseku než průměrná F0, přestože 90, respektive 102 sekund je přibližně pětinásobek hodnoty popsané výše jmenovanými autory.



Obr. 8.16. Postupná stabilizace průměrné F0 a základní hladiny u mluvčích M2 a M3. Osa pro průměr a ZH (v hertzech) šedě nalevo, osa pro průměr v půltónech (vzhledem k 1 Hz) černě napravo; obě osy zobrazují srovnatelný rozsah. Svislé přerušované čáry ukazují, jak počet replik u konkrétního mluvčího odpovídá trvání.

Porovnáme-li tyto výsledky s obrázkem 8.3 (ukazatele variability), zjistíme, že se jedná o očekávatelnou tendenci. U mluvčího M2 jsme pozorovali velmi nízkou směrodatnou odchylku, což je v souladu s rychlou stabilizací průměru F0. Výhoda základní hladiny oproti průměru se u tohoto mluvčího neprojevuje. Mluvčí M3 měl naopak směrodatnou odchylku F0 vysokou a v jeho případě (stejně jako v případě ostatních mluvčích, pro něž zde průběh stabilizace nezobrazujeme) se dá hovořit o výrazné robustnosti základní hladiny vůči variabilitě základní frekvence: ZH se stabilizuje výrazně rychleji než průměr. Vzhledem k často omezenému trvání forenzních nahrávek se proto základní hladina jeví jako o to vhodnější.

9 Závěr

Tato diplomová práce navazuje na současný stav bádání na poli forenzní fonetiky, které se odehrává ve znamení hledání idiosynkratických prvků v řečovém chování mluvčích. Toto bádání bývá zaměřené na spektrální nebo temporální charakteristiky řeči (nově tyto výzkumy prováděné na českém materiálu kompiluje monografie Fonetická identifikace mluvčího, 2014), jež lze analyzovat informovaným poslechem doplněným akustickou analýzou digitálních signálů. Pro účely naší práce jsme z nejčastěji zkoumaných domén vybrali základní frekvenci.

V úvodní části práce jsme vysvětlili okolnosti samotného vzniku hlasu: popsali jsme fyziologii hrtanu a mechaniku fonace. Od produkce jsme dále pokračovali k akustickým aspektům základní frekvence, odkud jsme se dále přesunuli ke kapitole o percepci a intonaci. Následující oddíl se již úžeji zaměřil na forenzní situaci, rozbory nahrávek, které je možné provádět a jaké závěry z nich lze vyvozovat. Poté jsme charakterizovali střední hlasovou frekvenci a vedle ní si v přípravě na empirickou část jako alternativní deskriptor představili robustní ukazatel základní hladiny, neutrální úrovně, na niž se podle autorů modulační teorie F0 daného mluvčího neustále vrací. Následoval oddíl, kterým jsme stručně představili druhy akustických analýz, zejména statických ukazatelů, jež jsme zamýšleli použít dále.

V experimentální části práce jsme prozkoumali, které ukazatele jsou pro statistický popis hodnot F0 výhodnější pro rozlišení mluvčích a mají větší výpovědní hodnotu než jiné, a to na dvou druzích materiálu získaného od osmi mužských mluvčích ve věkovém rozmezí 20-30 let: na spontánním rozhovoru a na čtených zprávách. Nejprve jsme se věnovali středním hodnotám F0 spontánního materiálu, průměru a mediánu, které sice neprojevily příliš velkou schopnost vystihnout rozdíly mezi mluvčími každý zvlášť, ale posuzovány dohromady (např. medián na pozadí průměru) a v kombinaci se základní hladinou dokázaly prozradit ty mluvčí, kteří se od ostatních lišili určitým idiosynkratickým návykem.

Základní hladina se projevila jako užitečný ukazatel v několika případech. Zaprvé reflektovala přítomnost třepené fonace, dokonce ji identifikovala lépe nežli poslech. Zadruhé se potvrdilo, co předpověděli již autoři tohoto ukazatele, a sice že je ZH odolná vůči vlivu různých mluvních stylů, a to lépe než průměr a medián. Zatřetí se ukázala její výhoda (opět oproti průměru) spočívající v rychlosti její stabilizace jako deskriptoru. Základní hladina je

tedy slibný ukazatel, který by při forenzní analýze měl dostat přednostní pozornost a od nějž by se poté mělo odvíjet hodnocení informací od ostatních deskriptorů.

Když jsme se dále zaměřili na ukazatele variability a zejména pak jejich srovnání napříč mluvními styly, ukázalo se, že toho o řečovém chování mluvčích napovídají více. Směrodatná odchylka pouze spolehlivě korespondovala s tvarem distribucí hodnot kolem průměru, zato rozpětí mezi 10. a 90. percentilem se (podobně jako medián ve srovnání s průměrem) projevilo jako užitečnější než obyčejné rozpětí. Při srovnání obou mluvních stylů tento ukazatel potvrdil vyšší variabilitu pro spontánní styl a spolehlivě korespondoval s poznatky, které jsme o mluvčích získali z histogramů F0. Konfrontace našich výsledků s dlouhodobými distribucemi F0 u všech mluvčích se obecně projevilo jako přínosné pro identifikaci mluvčích se skutečně idiosynkratickým chováním. Dále se osvědčily krabicové grafy, které jsou určitým integrovaným zobrazením ukazatelů střední hodnoty, variability a částečně i tvaru distribuce v jednom.

Jedním ze závěrů srovnání spontánního a čteného mluvního stylu bylo, že někteří mluvčí mají výrazně odlišné řečové návyky při čtení než při spontánní mluvě. Z toho vyplývá obecné doporučení pro forenzní praxi, v níž se někdy v rámci pořízení srovnávacích nahrávek sahá ke čtenému materiálu, aby byla zajištěna textová identita se spornou nahrávkou. V každém případě je ale třeba pro tuto chvíli zaměřit se na srovnání nahrávek stejného mluvního stylu. Proto by se měly budoucí snahy upnout dvojím směrem: jednak k nalezení dalšího ukazatele kromě základní hladiny, který by byl robustní vůči odlišnosti těchto dvou stylů, jednak k vynalezení způsobu, jak pořídit nahrávku srovnatelnou s reálnou spornou, ale to je spíše úkol pro psychology.

Po uvážení všech poznatků získaných z analýz deskriptorů základní frekvence můžeme uzavřít tuto práci s tím, že hledání idiosynkratického chování viditelného na F0 je obtížnější úkol, než když se o totéž fonetičtí badatelé snaží například u vokalických formantů. Jako doplňková analýza však zkoumání základní frekvence smysl má, zvláště když se provádí informovaně. Cílem této práce bylo tuto informovanost podpořit, aby v budoucnu byla interpretace ukazatelů přímočařejší a úspornější. Úkolem do budoucna nadále zůstává nacházet další silně idiosynkratické jevy (jako byla v této práci třepená fonace), které se napříč těmito ukazateli ve forenzním kontextu spolehlivě projeví.

10 Bibliografie

Beck, Janet M. (2005). Perceptual analysis of voice quality: the place of Vocal Profile Analysis. In: *A Figure of Speech: a Festchrift for John Laver*. London: Laurence Erlbaum, 285-322.

Boersma, P. (1993). Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound. *Proceedings of the Institute of Phonetic Sciences, 1193*, 97-110.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer (Verze 5.3.80). Staženo z <http://www.praat.org>.

Bořil, T. a Weingartová, L. (2014). Rozhodování a statistika. In: Skarnitzl, R. (Ed.), *Fonetická identifikace mluvčího*, 116-135. Praha: FF UK.

Braun, A. (1995). Fundamental frequency – How speaker-specific is it? *BEIPHOL 64, Studies in Forensic Phonetics*, 9–23.

Brown, W. S., Murry, T. a Michel, J. F. (1989). Vocal Jitter in Young Adult and Aged Female voices. *Journal of Voice, 3,* 113-119.

Coleman, R. a Markham, I. (1991). Normal Variations of Habitual Pitch. *Journal of Voice, 5,* 173-177.

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.

Gfroerer, S. a Wagner, I. (1995). Fundamental frequency in forensic speech samples. *BEIPHOL 64, Studies in Forensic Phonetics*, 41–48.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation.* Cambridge: CUP.

Hirose, H. (2010). Investigating the Physiology of Laryngeal Structures. In: W. J. Hardcastle, J. Laver a F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences,* 130-152. Oxford: Wiley-Blackwell.

Hirson, A., French, P. & Howard, D. (1995). Speech fundamental frequency over the telephone and face-to-face: some implications for forensic phonetics. In J. Windsor Lewis

(Ed.), *Studies in General and English Phonetics: Essays in Honour of Professor J.D. O'Connor*, 230–240. London: Routledge.

Hollien, H., Hollien, P. A. & de Jong, G. (1997). Effects of three parameters on speaking fundamental frequency. *Journal of the Acoustical Society of America*, *102*, 2984–2992.

Hollien, H. a Schwartz, R. (2000). Aural-Perceptual Speaker Identification, *Forensic Linguistics, 7*, 199-211.

Hudson, T., de Jong, G., McDougall, K., Harrison, P. & Nolan, F. (2007). F0 statistics for 100 young male speakers of Standard Southern British English. *Proceedings of the 16th ICPhS*, 1809–1812.

Jessen, M., Köster, O. & Gfroerer, S. (2005). Influence of vocal effort on average and variability of fundamental frequency. *International Journal of Speech, Language and the Law*, *12*, 174–213.

Kreiman, J. & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Malden, MA: Wiley-Blackwell.

Lindh, J. (2006). Preliminary descriptive F0-statistics for young male speakers. *Lund University Working Papers*, *52*, 89–92.

Lindh, J. & Eriksson, A. (2007). Robustness of long time measures of fundamental Frequency. *Proceedings of Interspeech 2007*, 2025–2028.

Nishio, M. a Niimi, S. (2008). Changes in Speaking Fundamental Frequency Characteristics with Aging. *Pholia Foniatr Logop., 60,* 120-127.

Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. *Proceedings of the 15th ICPhS*, 771–774.

Noteboom, S. (1997). The Prosody of Speech: Melody and Rhythm. In: Hardcastle, W. J. and Laver, J. (Eds.), *The Handbook of Phonetic Sciences*, pp. 640-673. Oxford: Blackwell.

Ohala, J. J. a Eukel, B. W. (1987). Explaining the Intrinsic Pitch of Vowels, In: R. Chanon a L. Shockey (Eds.), *In honor of Ilse Lehiste*, 207-215. Dodrecht: Foris.

Rietveld, A.C.M. & Gussenhoven, C. (1985). On the relation between pitch excursion size and pitch prominence*. Journal of Phonetics, 13*, 299-308.

Roach, P. (2006). *English Phonetics and Phonology: A Practical Course*. Cambridge: CUP.

Seikel, A. J., King, D.W a Drumright, D. G. (2010). *Anatomy and Physiology for Speech, Language, and Hearing*. New York: Thomson Delmar Learning.

Shipp, T. a McGlone, R. (1971). Laryngeal dynamics associated with voice frequency range. *Journal of Speech and Hearing Research, 14,* 761-768.

Skarnitzl, R. (2011). Znělostní kontrast nejen v češtině**.** Praha: Nakladatelství Epocha.

Skarnitzl, R. (2014). Forenzní fonetika. In: Skarnitzl, R. (Ed.), *Fonetická identifikace mluvčího*, 11-20. Praha: FF UK.

Skarnitzl, R. a Hývlová, D. (2014). Statistický popis hodnot základní frekvence. In: Skarnitzl, R. (Ed.), *Fonetická identifikace mluvčího*, 49-64. Praha: FF UK.

Skarnitzl, R., Lazárková, D., Nechanský, T., Šturm, P. (2014). Vokalické formanty. In: Skarnitzl, R. (Ed.), *Fonetická identifikace mluvčího*, 21-48. Praha: FF UK.

Skarnitzl, R. a Vaňková, J. (2015). Presenting the Population Statistics of Common Czech: Preliminary F0 Results. Prezentováno na *IAFPA 2015*, Leiden.

Svobodová, M. a Voříšek, L. (2014). Identifikace mluvčích z pohledu autentické kriminalistické praxe v České republice. In: Skarnitzl, R. (Ed.), *Fonetická identifikace mluvčího*, 136-144. Praha: FF UK.

t‘ Hart, J., Collier, R. a Cohen, A. (1990). A Perceptual Study of Intonation: *An experimental-phonetic approach to speech melody*. Cambridge: CUP.

Titze, I. R. (1986). Mean intraglottal pressure in vocal fold oscillation. *Journal of Phonetics, 14,* 359-364.

Titze, I. R. (1994). *Principles of Voice Production*. Englewood Cliffs: Prentice Hall, 191-217.

Traunmüller, H. (1994). Conventional, biological, and environmental factors in speech communication: A modulation theory. *Phonetica*, *51*, 170–183.

Traunmüller, H. a Eriksson, A. (1995). The frequency range of the voice fundamental in the speech of male and female adults. Staženo z <http://www2.ling.su.se/staff/hartmut/f0_m&f.pdf>

Vaissière, J. (2008). Perception of Intonation. In: *The Handbook of Speech Perception*, 236-263, Blackwell Publishing Ltd.

Volín, J. (2007a). Data volume requirements for reliable F0 normalization. In R. Vích (Ed.), *17th Czech-German Workshop - Speech Processing*, 62–67. Praha: AV ČR.

Volín, J. (2007b). *Statistické metody ve fonetickém výzkumu*. Praha: Nakladatelství Epocha.

Weingartová, L., Bořil, T. a Vaňková J. (2014). Spektrální sklon. In: Skarnitzl, R. (Ed.), *Fonetická identifikace mluvčího*, 77-94. Praha: FF UK.

Zraick, R. I., Gentry, M. A., Smith-Olinde, L. & Gregg, B. A. (2006). The effect of speaking context on elicitation of habitual pitch. *Journal of Voice*, *20*, 545–554.